Studying Correlations between Genetic Variation and Test Score Gaps

Edited by

Gregory Connor

John Fuerst

Studying Correlations between Genetic Variation and Test Score Gaps

Studying Correlations between Genetic Variation and Test Score Gaps

Edited by

Gregory Connor and John G.R. Fuerst

Cambridge
Scholars
Publishing

Studying Correlations between Genetic Variation and Test Score Gaps

Edited by Gregory Connor and John G.R. Fuerst

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2024 by Gregory Connor and John G.R. Fuerst and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-7566-7
ISBN (13): 978-1-5275-7566-0

TABLE OF CONTENTS

Introduction 1
Gregory Connor and John G.R. Fuerst

Part 1: Social and political considerations in research on test score gaps

Chapter 1 10
Research on group differences in intelligence: A defense of free inquiry
Nathan Cofnas

Chapter 2 28
How stifling debate around race, genes and IQ can do harm
Noah Carl

Chapter 3 41
Between-group differences in intelligence in the United States are >0% genetically caused: Five converging lines of evidence
Russell T. Warne

Part 2: Admixture regression analysis of race-related test score gaps

Chapter 4 70
Continental genetic ancestry source correlates with global cognitive ability score
Russell T. Warne

Chapter 5 86
Biogeographic ancestry, cognitive ability and socioeconomic outcomes
Emil O.W. Kirkegaard, Robert L. Williams, John G.R. Fuerst, and Gerhard Meisenberg

Chapter 6 115
Global ancestry and cognitive ability
Jordan Lasker, Brian J. Pesta, John G.R. Fuerst, and Emil O.W. Kirkegaard

Chapter 7 146
More research needed: There is a robust causal vs. confounding problem for intelligence-associated polygenic scores in context to admixed American populations
John G.R. Fuerst, Emil O.W. Kirkegaard, and Davide Piffer

Chapter 8 174
Linear and partially linear models of behavioral trait variation using admixture regression
Gregory Connor and John G.R. Fuerst

Chapter 9 203
Genetic ancestry and general cognitive ability in a sample of American youths
John G.R. Fuerst, Meng Hu, and Gregory Connor

Chapter 10 231
A genetically-informed test of the cognitive-colorism hypothesis
Vladimir Shibaev and John G.R. Fuerst

Chapter 11 260
Robustness analysis of African genetic ancestry in admixture regression models of cognitive test scores
John G.R. Fuerst

Part 3: Statistical differences in brain structure across biogeographic ancestries

Chapter 12 276
A genetic hypothesis for American race/ethnic differences in mean g: A reply to Warne (2021) with fifteen new empirical tests using the ABCD dataset
John G.R. Fuerst, Vladimir Shibaev, and Emil O.W. Kirkegaard

Chapter 13 333
A multimodal MRI-based predictor of intelligence and its relation to race/ethnicity
Emil O.W. Kirkegaard and John G.R. Fuerst

Supplemental materials 352

Contributors 353

INTRODUCTION

GREGORY CONNOR AND JOHN G.R. FUERST

In April 2003, the Human Genome Project announced the successful completion of its project to sequence the three billion base pairs of DNA in the human genome. This landmark achievement sparked an explosion in new genomic technologies, with widespread applications across multiple disciplines. Among a myriad of applications, these new genomic technologies allowed a major step forward in the analysis of genetic variation as a partial cause of intelligence test score gaps between racial and ethnic groups. This volume includes some of the most significant recent contributions to this important research topic. This brief introduction to the volume does not attempt to summarize all the papers herein; it merely provides a birds-eye view and highlights a few key findings shared among them.

The chapters in Part 1 of the book defend an open and uncensored approach to scientific inquiry. This approach is a widely accepted foundation for scientific research in most fields, but unusually controversial in the study of genetic variation and test score gaps. In chapters 1 and 2, Cofnas and Carl each address the question of whether, on balance, research on the possible genetic component of IQ gaps between racial groups should be suppressed. Both these authors employ a utilitarian framework, in this way evaluating whether suppression of scientific findings in this area has positive social value, against the alternative that scientific openness instead might be superior. They both acknowledge that there are potential positive and negative consequences of allowing race and IQ research. Each argues that, on balance, the standard practice of scientific openness is the wiser science policy choice.

Both authors document the extremely harsh, sometimes threatening or even violent, treatment of researchers whose published findings or public statements run contrary to the acceptable public narrative that IQ gaps are entirely environmentally caused. Carl notes that a utilitarian perspective is not universally shared by the opponents of scientific openness; many opponents treat biological identity between races as a “sacred value” and thereby exempt from utilitarian considerations:

> Why does this area of research incite such vitriolic indignation? A likely reason, as Winegard and Winegard (2015) argue, is that for a large number of academics in the West, the notion of biological sameness between groups (classes, sexes, races) has become what Tetlock (2003) calls a “sacred value” (and see Ginges et al., 2017). Sacred values possess at least two important properties. First, they are incommensurable with respect to instrumental values: no amount of sacred value can be traded off for any amount of an instrumental value. And second, proposals to accept such trade-offs are met not merely with rejection, but with moral outrage. Because arguments such as Wilson’s, Jensen’s and Murray’s clearly threaten the sacred value of biological sameness between groups, it is not enough simply to attack the arguments; the defenders of those arguments must be hounded, and their characters impeached. (Carl, Chapter 2 below).

Cofnas reflects on David Reich’s warning regarding likely future findings and their implications for current denialism regarding genetic causes of population trait differences:

> I have deep sympathy for the concern that genetic discoveries about differences among populations may be misused to justify racism. But it is precisely because of this sympathy that I am worried that people who deny the possibility of substantial biological differences among populations across a range of traits are digging themselves into an indefensible position, one that will not survive the onslaught of science. In the last couple of decades, most population geneticists have sought to avoid contradicting the orthodoxy. When asked about

> the possibility of biological differences among human populations, we have tended to obfuscate, making mathematical statements in the spirit of Richard Lewontin about the average difference between individuals from within any one population being around six times greater than the average difference between populations … but this carefully worded formulation is deliberately masking the possibility of substantial average differences in biological traits across populations. (Reich 2018, quoted in Cofnas, Chapter 1 below).

Only a few years have passed since Reich wrote those words, but his forecast seems already realized. New findings on genetic causes of test score gaps are statistically strong, robust to model specification, and have been replicated across multiple independent databases. Genetic variation as one source of race-related test score gaps is no longer just a hypothetical possibility, the existing evidence is now quite strong. This change in scientific understanding is due in part to the papers in Part 2 of this volume. This new reality adds urgency to Reich's call to abandon rigid denialism regarding genetic causes of trait variation across groups.

In Chapter 3, Warne lays out a strong empirical case that some test score gaps have partly genetic causes by reviewing five separate lines of supporting evidence. He laments that the politicization of funding opportunities and other politically motivated blockages on this line of research have impeded scientific progress; he issues a call for more open-minded and uncensored research:

> Combined, all five sources of data make a cohesive argument that group differences in intelligence are not fully environmentally caused. I call on psychologists to have an open mind and to investigate the evidence for themselves, starting with the sources I have cited in this article. I also encourage social scientists to make research contributions that can address this question… Controversy thrives in ignorance; it is time to move beyond the basic question of whether genes have an impact on intergroup mean IQs so that scientists can investigate productive questions of the magnitude of this influence and how to accommodate differences in a diverse society. (Warne, Chapter 3 below)

Parts 2 and 3 of the volume are predicated on the open scientific approach advocated in Part 1. Part 2 contains the core scientific material of the volume, with a strong focus on admixture-based analysis of test score gaps.

The distinction between social race, a socially defined identity, and biogeographical ancestry, a genetically determined characteristic, is a crucial component in all the papers in Part 2. Social race refers to an individual's socially defined membership of race or ethnic groups such as Black, white, Hispanic, Native American, Chinese, etc.. This socially defined identity is easily captured, for the vast majority of US individuals, by simply asking each individual to choose one or more from a list of race/ethnicity categories. Biogeographic ancestry (hereafter BGA) denotes an individual's proportional genetic ancestry across a set of continental populations such as African, European, Amerindian, and Asian. Prior to the genetic revolution of the last two decades, these BGA proportions could only be measured very imprecisely. Nowadays, sophisticated new cluster algorithms applied to genotyped DNA samples can identify and accurately measure BGA proportions for most US individuals.

The genotype-based technology to separate SIRE from BGA has sparked off a small scientific revolution in the application of genetic methods to racial identity-related trait variation. Admixture regression involves least-squares regression of a trait on a set of BGA proportions and a set of SIRE variables simultaneously, along with other relevant variables. Admixture regression greatly improves the ability to separate environmental and genetic causes of ethnic and racial group differences in an array of medical, anthropological, and behavioral traits. Admixture regression analysis has been used over the last two decades by genetic epidemiologists and other researchers to study race and ethnicity related differences in alcohol dependence (Zuo et al., 2009), height (Becker et al., 2011), asthma risk (Flores et al., 2012), cardiovascular disease (Bidulescu et al., 2014), sleep depth (Halder et al., 2015), cigarette smoking behavior (Choquet, Yin, and Jorgenson, 2021), metabolomics (Mehanna et al., 2022), cancer (Rhead et al., 2022), and diabetes (Parcha et al., 2023). The papers in Part 2 are key contributions to a powerful wave of new research applying modern admixture regression methodologies to the investigation of group differences in intelligence. Admixture analysis had long been acknowledged as a potentially powerful method for identifying partly genetic causes of racial and ethnic group differences in intelligence (Shockley, 1973; Nisbett, 2009). Early studies from the 1960s and 1970s had to rely on ancestry estimates based on blood types. These estimates proved unreliable (Reed, 1997) and so more reliable analysis had to wait until admixture proportions could be measured accurately.

In Chapter 4, Warne gives a lucid introduction to the use of admixture regression methods for measuring the contribution of genetic variation to race-related test score gaps. The empirical analysis in Warne's paper is impaired by the small sample size of his database. The use of on-line intelligence tests and self-reported admixture measurement, rather than professionally administered intelligence tests and lab-linked admixture measurement, are other data-related limitations of his study. These data limitations restrict the depth of analysis that it is attainable relative to other recent admixture studies based on large, professionally constructed databases. Despite its data limitations, the paper is a highly recommended introduction to the admixture regression literature. Interestingly, the obvious concern about the replicability of the paper's small-sample findings did not turn out to be a problem – subsequent papers using large, professionally constructed databases reach the same empirical conclusions as Warne did in his modest, exploratory empirical study.

The choices of SIRE categories to include/exclude from the estimation sample, and the choice of BGA clusters, depend upon the available data set and the particularities of the research question. The choices made across the chapters in Part 2 are quite diverse; see Table 1 below and the individual chapters for greater detail.

Table 1: Self-Identified Race and Ethnicity (SIRE) and Biogeographic Ancestry (BGA) Categories Used in the Chapters in Part 2

Chapter	SIRE Categories in Estimation Sample	BGA Categories
4	Black and Hispanic	European, African, Other
5	Black, white, Native American, Asian, Hispanic, Pacific Islander, Multiple-SIRE, and Other SIRE	European, African, Amerindian, East Asian, Central Asian, Oceanian
6	Non-Hispanic Black, non-Hispanic white, and non-Hispanic Black-white biracial	European and African
7	Hispanic, non-Hispanic white, non-Hispanic Black, and non-Hispanic Black-white biracial	European, African, Amerindian
8	Black, white, Hispanic, Native American, East Asian, South Asian, and Other including Multiple-SIRE	European, African, Amerindian, East Asian, and South Asian
9	Black, white, Hispanic, Native American, East Asian, South Asian, and Other including Multiple-SIRE	European, African, Amerindian, East Asian, and South Asian
10	Black, Hispanic, Native American, and Other	European, Amerindian, African, East Asian, and South Asian
11	Black, white, Hispanic, Native American, and Other	European, African, Amerindian, East Asian, and South Asian

There are various other model specification choices that differ across the chapters. One is the choice between including Black-white biracial SIRE as a separate dummy variable (e.g., Chapter 6) versus assigning Black-white biracials equal ½-½ fractional weightings on the Black and white SIRE dummy variables (e.g., Chapter 9). Black-white biracial is one of the few multiple-SIRE cases where adding a dedicated dummy variable is practical. It is not feasible to add separate dummy variables for all possible multiple-SIRE combinations since some combinations have very few observations; it is necessary either to drop multiple-SIRE individuals from the estimation sample or to use fractional dummy weightings for them. Another specification choice involves choosing whether Hispanic ethnicity should be included separately as a zero-one dummy in the regression or treated equivalently to other SIRE choices with fractional weighting on Hispanic identity and other racial identities; see Chapter 5 which tries both specification choices and compares them.

A difficult specification choice concerns whether to include socio-economic status (SES) as a separate explanatory variable in the regression. Genetic variation causes SES differences across families, so including SES as an explanatory variable will tend to understate total genetic influence of BGA on test scores. On the other hand, SES can directly influence test scores due to environmental effects arising from a less privileged environment, so excluding it may give rise to missing variable biases in the other estimated coefficients. In many of the chapters the admixture regression is estimated both with and without SES, and the reader is left to choose which seems more informative.

The standard regression estimation algorithm does not allow a set of proportion variables which sum to one for every observation, nor a set of dummy variables such that each observation has unit summed exposure, since together with the intercept such a set of explanatory variables generates a linear singularity in the regression matrix. The standard procedure is to delete one of the proportion variables and/or one of the dummy variables to be used as a base case; the remaining estimated coefficients then capture differences relative to that base case variable. Most of the papers in Part 2 delete the European BGA variable and the white SIRE variable from the admixture regressions since these have the most observations in the sample data and thereby are suitable base cases.

Reviewing all the chapters in Part 2, perhaps the strongest empirical finding is that African ancestry has a substantial and statistically significant negative impact on cognitive ability test scores (see Ch. 4, Table 2; Ch. 5, Table 7, model 2; Ch. 7, Table 2; Ch. 8, Table 2; Ch. 9, Table 5; Ch. 11, Tables 1- 6). Black SIRE and African biogeographic ancestry overlap broadly in this US data, hence when SIRE variables are used alone in a regression (excluding BGA variables) Black SIRE has a negative relationship to test scores. However, this Black SIRE impact on test scores appears to be due mostly to its proxy role for African biogeographic ancestry (compare Chapter 5, Table 7, model 1 versus model 2; or Ch. 8, Table 1 versus Table 2). This does not imply that the direct effect of Black SIRE on test scores is zero, only that it seems too small to separately identify in these moderately-large databases. There is also some evidence, not as uniformly powerful across the chapters, that Amerindian BGA has a negative statistical impact and East Asian BGA possibly has a positive statistical impact on test scores (using European BGA as the base case).

The papers in the volume interpret the relation between biogeographic ancestry and cognitive ability as evidence in support of a partial genetic hypothesis for group differences in test scores, but none of them taken individually control for every possible confounding factor. Rather, following conventional scientific practices, the papers build on one another, addressing weaknesses in prior research, noting current limitations, and proposing directions for further research. As in the case of other behavioral genetic research (e.g., adoption studies, twin studies, and genome-wide association studies) no single analysis is dispositive; the strength of the evidence comes from the accumulated results based on multiple, overlapping research designs.

None of the papers fully answer the question as to what caused the observed differences in genetic variation across ancestries. Some of the papers use evolutionary theory and the history of human development to give theoretical reasons for justifying the tested hypothesis that racial test score gaps have partially genetic causes; see for instance chapters 5 and 6. This is an aspect of the research area that may see substantial progress over coming decades.

Genetic variation across ancestries as a source of trait variation implies that there are genetic variants related to the trait and that their population frequency differs across ancestries. It is not necessary to identify these specific genetic variants to perform admixture regression; the existence of such genetic variants is implied by an observed relationship between admixture and trait. Kirkegaard, et al. (Chapter 5) call this *global admixture analysis*, and they differentiate this from *admixture mapping* in which the researcher attempts to identify specific alleles (or tagged allelles serving as proxies) that covary in frequency with the trait and have measurably different frequencies across one or more BGA's. The papers in this volume mostly involve global admixture, but Chapters 6 and 12 take initial steps toward admixture mapping of the BGA-test score relationship. This involves employing polygenic indices for intelligence or educational attainment (constructed by other researchers) and examining the relationship of their index scores to BGA proportions.

Part 3 of the book looks toward the scientific future. It contains two papers that take a natural next step in this research area: exploring whether test score gaps can be linked to statistical differences in the physical characteristics of the human brain across biogeographic ancestries. Chapter 12 (Fuerst, Shibaev, and Kirkegaard) is a monumental study which tests a wide array of empirical hypotheses about genetic variation, brain size and structure, polygenic indices, and intelligence test scores. Chapter 13 (Kirkegaard and Fuerst) relies on the magnetic resonance image data included in the Adolescent Brain and Cognitive Development (ABCD) database and uses the data to explore brain structure differences across biogeographic ancestries. The paper applies an elastic net statistical framework to build a prediction model of intelligence based on high-dimensional multimodel MRI data. They show that an MRI-based intelligence prediction model estimated using the white-only subsample has substantial predictive power for the Black-only subsample, and that this MRI-based model explains 37% of the observed test score gap between the Black and white subsamples even though it is estimated on the white-only subsample. This is clear preliminary evidence for physically-observable genetically-mediated causes of the Black-white test score gap.

Over fifty years ago, Arthur Jensen (1969) challenged the prevailing orthodoxy that test score gaps between racial groups can be ascribed solely to environmental causes. The environment-only theory which Jensen challenged has maintained its hegemony as the only publicly acceptable theory. Jensen made his case using indirect methods of statistical inference since genomic technology at that time did not allow more direct analysis of genetic data. With the advent of the genomic revolution, quantitative genetic methods including admixture regression have revolutionized research on the causes of test score gaps. The papers in this volume provide a powerful body of evidence in support of Jensen's fifty-year-old claim. None of the papers on its own is dispositive but taken together they constitute a powerful new body of evidence contributing to a better understanding of these recalcitrant test score gaps. With this new genetic evidence, the environment-only theory is less scientifically plausible.

Acknowledgements: All the papers in this volume appeared previously in other publication outlets; we thank the authors and previous publishers for permission to reprint them here. The original publication citations are included in the reference list below.

References

Becker, N. S., Verdu, P., Froment, A., Le Bomin, S., Pagezy, H., Bahuchet, S., and Heyer, E. (2011). Indirect evidence for the genetic determination of short stature in African Pygmies. *American journal of physical anthropology,* 145(3), 390-401.

Bidulescu, A., Choudhry, S., Musani, S. K., Buxbaum, S. G., Liu, J., Rotimi, C. N., … and Gibbons, G. H. (2014). Associations of adiponectin with individual European ancestry in African Americans: the Jackson Heart Study. *Frontiers in genetics*, 5, 22.

Carl, N. (2018) How stifling debate around race, genes and IQ can do harm*, Evolutionary Psychological Science*, 4: 399-407.

Cofnas, N. (2020) Research on group differences in intelligence: A defense of free inquiry, *Philosophical Psychology*, 33:1, 125-147.

Colman, A. M. (2016). Race differences in IQ: Hans Eysenck's contribution to the debate in the light of subsequent research. *Personality and Individual Differences*, 103, 182-189.

Connor, G., and J.G. Fuerst (2022) Linear and partially linear models of behavioral trait variation using admixture regression. In R. Lynn (Ed.). *Intelligence, Race and Sex: Some Controversial Issues: A Tribute to Helmuth Nyborg at 85*. Arktos Publishers.

Choquet, H., Yin, J., and Jorgenson, E. (2021). Cigarette smoking behaviors and the importance of ethnicity and genetic ancestry. *Translational psychiatry*, 11(1), 1-10.

Flores, C., Ma, S. F., Pino-Yanes, M., Wade, M. S., Pérez-Méndez, L., Kittles, R. A., … and Garcia, J. G. (2012). African ancestry is associated with asthma risk in African Americans. *PloS one*, 7(1), e26807.

Fuerst, J. G. (2021). Robustness Analysis of African Genetic Ancestry in Admixture Regression Models of Cognitive Test Scores. *Mankind Quarterly*, 62(2).

Fuerst, J.G.R., E.O.W. Kirkegaard and D. Piffer (2021). More Research Needed: There is a Robust Causal vs. Confounding Problem for Intelligence-associated Polygenic Scores in Context to Admixed American Populations. *Mankind Quarterly* 62(1):151-185.

Fuerst, J.G., E.O. Kirkegaard and V. Shibaev (2023). A Genetic Hypothesis for American Race/Ethnic Differences in Mean g: A Reply to Warne (2021) with Fifteen New Empirical Tests Using the ABCD Dataset. *Mankind Quarterly* 63(4):527-600.

Fuerst, J.G., M. Hu, and G. Connor (2021) Genetic ancestry and general cognitive ability in a sample of American youths, *Mankind Quarterly*, *62:*1, 186–216

Gottfredson, L. S. (2007). Applying double standards to "divisive" ideas. *Perspectives on Psychological Science, 2*(2), 216-220.

Halder, I., Matthews, K. A., Buysse, D. J., Strollo, P. J., Causer, V., Reis, S. E., and Hall, M. H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep*, *38*(8), 1185-1193.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.

Kirkegaard, E.O.W., & Fuerst, J.G.R. (2023). A multimodal MRI-based predictor of intelligence and its relation to socio-demographic characteristics. *Mankind Quarterly,* 63: 374-397.

Kirkegaard, E.O., R.L. Williams, J.G. Fuerst and G. Meisenberg. (2019). Biogeographic Ancestry, Cognitive Ability and Socioeconomic Outcomes. *Psych*, *1*(1), 1-25.

Lasker, J., B.J. Pesta, J.G. Fuerst and E.O. Kirkegaard (2019), Global ancestry and cognitive ability, *Psych* 1(1): 431-459.

Mehanna, M., McDonough, C. W., Smith, S. M., Gong, Y., Gums, J. G., Chapman, A. B., ... and Cooper-DeHoff, R. M. (2022). Influence of Genetic West African Ancestry on Metabolomics among Hypertensive Patients. *Metabolites*, *12*(9), 783.

Meyer, M. N., Appelbaum, P. S., Benjamin, D. J., Callier, S. L., Comfort, N., Conley, D., ... and Parens, E. (2023). Wrestling with Social and Behavioral Genomics: Risks, Potential Benefits, and Ethical Responsibility. *Hastings Center Report*, *53*, S2-S49.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. WW Norton and Company.

Parcha, V., Heindl, B., Kalra, R., Bress, A., Rao, S., Pandey, A., ... and Arora, P. (2022). Genetic European ancestry and incident diabetes in Black individuals: Insights from the SPRINT trial. *Circulation: Genomic and Precision Medicine*, *15*(1), e003468.

Reed, T. E. (1997). " The genetic hypothesis": It was not tested but it could have been. *American Psychologist*, 52(1), 77-78.

Rhead, B., Hein, D., Pouliot, Y., Guinney, J., De La Vega, F., and Sanford, N. N. (2022). Genetic ancestry differences in tumor mutation in early and average-onset colorectal cancer.

Reich, D. (2018). *Who we are and how we got here: Ancient DNA and the new science of the human past*. Pantheon Books, New York.

Shibaev, V. and J.R. Fuerst (2023). A Genetically Informed Test of the Cognitive-Colorism Hypothesis. *Natural Systems of Mind*, *3*(1).

Shockley, W. (1973). Variance of Caucasian admixture in Negro populations, pigmentation variability, and IQ. In *Proceeding of the National Academy of Sciences of the United States of America,* 70(7), 2180-2180.

Templeton, A.R. The genetic and evolutionary significance of human races. In *Race and Intelligence: Separating Science from Myth*; Fish, J.M., Ed.; Routledge: Abingdon, UK, 2001.

Warne, R. T. (2020). Continental Genetic Ancestry Source Correlates with Global Cognitive Ability Score. *Mankind Quarterly*, 60(3), 400–422.

Warne, R.T. (2021) Between-group differences in intelligence in the United States are >0% genetically caused: Five converging lines of evidence, *The American Journal of Psychology*, 134: 480-501.

Zuo, L., Luo, X., Listman, J. B., Kranzler, H. R., Wang, S., Anton, R. F., ... and Gelernter, J. (2009). Population admixture modulates risk for alcohol dependence. *Human genetics*, 125(5), 605-613.

PART 1:

SOCIAL AND POLITICAL CONSIDERATIONS IN RESEARCH ON TEST SCORE GAPS

CHAPTER 1

RESEARCH ON GROUP DIFFERENCES IN INTELLIGENCE: A DEFENSE OF FREE INQUIRY[1]

NATHAN COFNAS

Abstract

In a very short time, it is likely that we will identify many of the genetic variants underlying individual differences in intelligence. We should be prepared for the possibility that these variants are not distributed identically among all geographic populations, and that this explains some of the phenotypic differences in measured intelligence among groups. However, some philosophers and scientists believe that we should refrain from conducting research that might demonstrate the (partly) genetic origin of group differences in IQ. Many scholars view academic interest in this topic as inherently morally suspect or even racist. The majority of philosophers and social scientists take it for granted that *all* population differences in intelligence are due to environmental factors. I argue that the widespread practice of ignoring or rejecting research on intelligence differences can have unintended negative consequences. Social policies predicated on environmentalist theories of group differences may fail to achieve their aims. Large swaths of academic work in both the humanities and social sciences assume the truth of environmentalism and are vulnerable to being undermined. We have failed to work through the moral implications of group differences to prepare for the possibility that they will be shown to exist.

Keywords: hereditarianism, environmentalism, utilitarianism, intrinsic value of truth

1. Introduction

The idea that all human groups have, on average, the same potential to develop all psychological traits and abilities is an empirically testable claim, but in some important ways it is unlike most other empirical claims investigated by scientists. Many scientists, philosophers, and even laypeople hold that, for a variety of reasons, we are morally obligated to affirm this proposition, or refrain from conducting research that might undermine it (e.g., Block & Dworkin, 1974; Chomsky, 1976, 1988; Dummett, 1981; Gardner, 2001; Horgan, 2013; Kourany, 2016; Sternberg, 2005). Prominent researchers who publicly endorse hereditarianism about group differences in intelligence[2] have been condemned as immoral, fired from their positions, and physically threatened (examples are documented in Ceci & Williams, 2009; Cofnas, 2016; Gottfredson, 2010; Sesardic, 2005). Many scholars assume that there is no legitimate reason to be interested in group differences. According to Rose (2009), the fact that psychometricians focus on the IQ gap between some groups (e.g.,

---

[1] Editors' note: We have reformatted the original paper for this collection.

[2] Practically all informed scholars are, to varying degrees, hereditarians about *individual* differences in intelligence (Neisser et al., 1996). The controversy centers on the role of genes in *group* differences. Hereditarians about group differences believe that genes play a nontrivial role in these differences—environmentalists, that genes play no significant role.

Blacks vs. Whites) rather than others (e.g., north vs. south Welsh) "calls into question the motivation behind looking for such specific group differences in intelligence, sheds doubt on whether such research is well-founded, and begs whether answers could possibly be put to good use" (p. 786). He concludes that "in a society in which racism and sexism were absent, the questions of whether whites or men are more or less intelligent than blacks or women would not merely be meaningless—they would not even be asked" (p. 788). Virtually no scientific hypothesis besides race and sex differences is rejected in this way.

The fact that the majority of genetic variation in our species exists within rather than between races, as Lewontin (1972) discovered, does not rule out the possibility of significant race differences in psychological traits like intelligence. Lewontin argued that, because there is more genetic variation within than between races, racial classifications do not correspond to any (genetically based) population structure. That implies that there can be no genetically based racial classification, thus there can be no genetically based racial differences. This argument overlooks the fact that racial classification can be based on small differences in gene frequencies at *multiple* polymorphic loci. Looking at a single locus gives us almost no taxonomically useful information, but looking simultaneously at multiple loci allows us to classify people into ethnic groups with accuracy that can approach 100% (Edwards, 2003; Tal, 2012). David Reich, who runs a leading genetics lab at Harvard University, recently acknowledged that Lewontin's argument has been used by geneticists to *"deliberately"* conceal the possibility of significant genetically based population differences:

> I have deep sympathy for the concern that genetic discoveries about differences among populations may be misused to justify racism. But it is precisely because of this sympathy that I am worried that people who deny the possibility of substantial biological differences among populations across a range of traits are digging themselves into an indefensible position, one that will not survive the onslaught of science. In the last couple of decades, most population geneticists have sought to avoid contradicting the orthodoxy. When asked about the possibility of biological differences among human populations, we have tended to obfuscate, making mathematical statements in the spirit of Richard Lewontin about the average difference between individuals from within any one population being around six times greater than the average difference between populations....But this carefully worded formulation is deliberately masking the possibility of substantial average differences in biological traits across populations. (Reich, 2018, p. 254)

Reich addresses another common argument used to reject the possibility of substantial race differences: even if there is genetically based variation "affecting cognition or behavior," these differences must be small because "so little time has passed since the separation of populations." He finds this argument untenable in the light of recent findings in genetics:

> The average time separation between pairs of human populations since they diverged from common ancestral populations, which is up to around fifty thousand years for some pairs of non-African populations, and up to two hundred thousand years or more for some pairs of sub-Saharan African populations, is far from negligible on the time scale of human evolution. If selection on height and infant head circumference can occur within a couple of thousand years, it seems a bad bet to argue that there cannot be similar average differences in cognitive or behavioral traits. Even if we do not yet know what the differences are, we should prepare our science and our society to be able to deal with the reality of differences instead of sticking our heads in the sand and pretending that differences cannot be discovered. (Reich, 2018, p. 258)

Section 2 explains why, from a scientific perspective, it is so difficult to determine the cause of group differences in intelligence. Although the genetic origin of race differences can only be definitively established when the fields of genetics and neuroscience are unified in a way that will not happen for several years, less definitive evidence that is currently available seems to implicate genes. Section 3 raises a question of values: should science be absolutely committed to the truth, or, as some philosophers and scientists believe, are topics such as race differences in intelligence just too dangerous to face openly? Section 4 argues that utilitarian considerations favor open inquiry into the cause of race differences. Section 5 argues that truth has value independent of its consequences, though this value should be balanced against the potential dangers of truth. Section 6 argues that philosophers should help devise an ethical way to respond to group differences in intelligence in anticipation of the possibility that they will be shown to exist.

2. The scientific controversy

The mere fact that populations differ in IQ tells us little or nothing about the *source* of the difference. The difference could be due to genes, but it could also be due to some environmental factor(s)—education, nutrition, cultural bias in IQ tests, and so on. How can we disentangle the effects of genes and environment? Behavioral geneticists have two principal methods for probing the role of genes in phenotypic differences: (a) behavioral genetic studies on adoptees and twins, and (b) genome-wide association studies.

The heritability of a trait—the proportion of the variance in a population explained by genetic differences—can be determined by adoption and twin studies. If adopted children grow up to resemble their *biological* rather than their *adoptive* parents, this implicates genes. If (genetically identical) monozygotic twins raised together grow up to resemble each other more than (genetically related but nonidentical) dizygotic twins raised together, this, too, implicates genes (on certain testable assumptions, such as equal environments or, in the case of adoption studies, random placement into adoptive environments). Adoption and twin studies show that, in the contemporary United States, the heritability of IQ in adulthood is around 60 to 80% (Plomin & Deary, 2015).

Even if IQ has high heritability within racial groups, this does not imply that race differences are genetic. We cannot infer *between*-group heritability from *within*-group heritability. Lewontin (1970, pp. 7–8) illustrates this point with the following example. Suppose we have a sack of genetically variable corn seeds, and we plant a handful on two plots of soil. Conditions within each plot are carefully controlled so that each plant receives the same nutrients, sunlight, and so on. Since there is no variation in the environmental factors that affect growth, differences in height within each plot will be due entirely to genes—height will be 100% heritable. But suppose one plot contains a low concentration of nitrates. The corn in the low-nitrates plot will grow shorter on average for purely environmental reasons. Within-group (i.e., within-plot) heritability will be 100%, while between-group heritability will be 0%.

Heritability studies cannot show definitively that race differences in intelligence have a genetic cause. It is always possible that there is some hidden environmental factor(s)—a so-called "X factor," analogous to nitrates in the corn example—that explains differences between but not within races. (An X factor that explains the black–white IQ gap would have to affect all Blacks equally in order to preserve high heritability and shift the population mean downwards without changing the variance.) Nevertheless, the high within-group heritability of IQ can be part of a *package* of evidence for between-group heritability. This point was made by Jensen (1969), who was widely misunderstood as naively inferring between- from within-group heritability (see Sesardic, 2005, pp. 128–138). Expounding on Jensen's argument, Flynn (1980) writes that "the probability of a genetic hypothesis will be much enhanced if, in addition to evidencing high [heritability] estimates, we find we can falsify literally every plausible environmental hypothesis [i.e., X factor] one by one" (p. 40; quoted in Sesardic, 2005, p. 136). The obvious candidate X factor that could explain race differences is, of course, *racism*. If racism lowers IQ, this could explain why the mean is shifted downwards for victimized groups. But, as Flynn argues, attributing differences to racism

> is simply an escape from hard thinking and hard research. Racism is not some magic force that operates without a chain of causality. Racism harms people because of its effects and when we list those effects, lack of confidence, low self-image, emasculation of the male, the welfare mother home, poverty, it seems absurd to claim that any one of them does not vary significantly within both black and white America. (Flynn, 1980, p. 60; quoted in Sesardic, 2005, pp. 141–142)

Now, after decades of intensive searching, the X factor remains elusive. The adult black–white IQ gap has remained stubbornly constant at approximately one standard deviation (15 IQ points) among cohorts born since around 1970 (Murray, 2007). Dickens and Flynn (2006) report that "Blacks gained 4 to 7 IQ points on non-Hispanic Whites between 1972 and 2002" (p. 913), but these gains appear to be among Blacks born before the early seventies. Dickens and Flynn (2006, Figure 3) indicate that, in 2002, the black–white IQ gap among 20-year-olds was approximately one standard deviation, or 15 points. Nisbett (2017) writes that "Dickens and Flynn found [the Black–White gap in IQ to be] around 9.5 points," but this is only the gap if we include *children* (as Nisbett confirmed in a personal communication). More recent evidence indicates that the gap has persisted or even widened. Frisby and Beaujean (2015, Table 8) find a black–white IQ gap of 1.16 standard deviations among a population-representative sample of adults used to norm the Wechsler Adult Intelligence Scale IV in 2007. Intensive interventions can raise IQ substantially during childhood, when the heritability of IQ is low. But despite some misleading claims about the success of early intervention programs, gains tend to dissolve by late adolescence or early adulthood (Baumeister & Bacharach, 2000; Lipsey, Farran, & Durkin, 2018; Protzko, 2015). Adoption by white families—one of the most extreme interventions possible—has virtually no effect on the IQ of black adoptees by adulthood. Black children adopted by middle- and upper-middle-class white families in Minnesota obtained IQ scores at age 17 that were roughly identical to the African American average. Adoptees with one black biological parent obtained IQ scores that were intermediate between the black and white means (Loehlin, 2000, Table 9.3).[3]

To reiterate, the high within-group heritability of IQ combined with the failure to find an environmental X factor to explain the IQ gap does not show decisively that race differences are genetic, because it is possible that an X factor will be discovered in the future. However, the environmentalist theory of race differences has not, by normal scientific standards, been an especially progressive research program (in the sense of Lakatos, 1970). Environmentalists never predicted that the black–white IQ gap would, after reaching one standard deviation, remain impervious to early education, adoption, massive improvements in the socioeconomic status of Blacks, and the (apparent) waning of overt racism and discrimination. Commenting 45 years ago on environmentalist theories that appeal to an X factor, Urbach (1974) noted that "any data in the world can be made consistent with any theory by invoking nameless and untested factors" (p. 134).

---

[3] The researchers who carried out the Minnesota Transracial Adoption Study argue that the "results...provide little or no conclusive evidence for genetic influences underlying racial differences in intelligence" (Waldman, Weinberg, & Scarr, 1994, p. 29). They give two main reasons for this conclusion. First, the mean IQ of Asian/Indian adoptees was lower than that of white adoptees, despite the fact that mean (East) Asian IQ is higher than that of Whites. This contradicts the hereditarian prediction that the rank ordering of the adoptees would reflect the rank ordering of their respective racial groups. Second, black children had worse early adoptive experiences than white children: Whites were placed with an adoptive family sooner and spent more time with them than Blacks (although half-black children were placed sooner than white children). Controlling for early experience explained a small proportion of the variance in IQ at age 17. But the fact remains that adoption did not have a significant effect on the IQ gap. Scarr (1998) later wrote: "The test performance of the Black/Black adoptees was not different from that of ordinary Black children reared by their own families in the same area of the country. My colleagues and I reported the data accurately and as fully as possible, and then tried to make the results palatable to environmentally committed colleagues. In retrospect, this was a mistake" (p. 230).

Let us turn to the second method for investigating the role of genes in development: genome-wide association studies (GWAS). Unlike heritability studies, GWAS can uncover *specific genetic variants*—or single-nucleotide polymorphisms (SNPs)—associated with IQ. In just the last couple years, GWAS has identified hundreds of such SNPs (Davies et al., 2018; Savage et al., 2018; Sniekers et al., 2017), which together explain around 11% of the variance in IQ (Allegrini et al., 2019).

If we find that the SNPs implicated in IQ are differentially distributed across racial groups, this would not necessarily imply that race differences in intelligence are genetic. SNPs might have different effects across races and environments due to gene–gene and gene–environment interactions. SNPs with no causal relation to intelligence can be genetically linked to SNPs that do have a causal relation in some populations but not others, so SNP–intelligence correlations may not always hold across races (Rosenberg, Edge, Pritchard, & Feldman, 2019). But if we find that many of the same SNPs predict intelligence *in different racial groups*, a risky prediction made by the hereditarian hypothesis will have passed a crucial test. Even then, however, GWAS will only establish a *correlation* between SNPs and IQ without revealing the causal chain linking SNP to phenotype. It would still be theoretically possible that these SNPs lead to differences in intelligence as a consequence of environmental factors (e.g., parenting effects) that can be manipulated so as to eliminate race differences. But if work on the genetics and neuroscience of intelligence becomes sufficiently advanced, it may soon become possible to give a convincing causal account of how specific SNPs affect brain structures that underlie intelligence (Haier, 2017). If we can give a biological account of how genes with different distributions lead to race differences, this would essentially constitute proof of hereditarianism. As of now, there is nothing that would indicate that it is particularly unlikely that race differences will turn out to have a substantial genetic component. If this possibility cannot be ruled out scientifically, we must face the ethical question of whether we *ought* to pursue the truth, whatever it may be.

3. Should science be absolutely committed to truth?

Truth was traditionally thought of—and is still often advertised—as the fundamental commitment of science (e.g., Dawkins, 2006). Some of the great heroes of science—most famously Galileo—are celebrated for resisting moral and religious authorities who feared that free inquiry would bring disaster. But the fact that the censors of previous eras were wrong does not mean that those of today cannot be right (Dummett, 1981, pp. 287–288). Sesardic (1992) (who ultimately defends free inquiry) observes that since the days when "truth was...regarded as the categorical imperative for scientists," we have "learned from some dramatic instances that there is no guarantee that...new knowledge [will] serve the good of mankind" (p. 129). Just as physics can produce an atom bomb with the potential to destroy us physically, perhaps social science can produce an equivalent—a discovery that would undermine the social order and destroy us just the same.

Many scientists claim to be absolutely committed to truth while at the same time advocating or tolerating varying degrees of suppression of controversial work. In his famous article, "The Responsibility of Intellectuals," Chomsky (1967) makes the bald statement that "it is the responsibility of intellectuals to speak the truth and to expose lies." Elsewhere, however, he says that scientists should not pursue the truth about group differences in intelligence because this could have dangerous social consequences (Chomsky, 1976, pp. 294–295). He says that while people "surely...differ in their biologically determined qualities," it is wrong to investigate an association between group membership and IQ because to do so is to indicate "that the answer to the question makes a difference; it does not, except to racists, sexists, and the like" (Chomsky, 1988, p. 164; quoted in Cofnas, 2016, p. 486). "To anyone not afflicted with these disorders, it is of zero interest whether the average value of IQ for some category of persons is such-and-such" (Chomsky, 1988, p. 164).

Cofnas (2016) documented several prominent scientists and philosophers opposing the study or dissemination of findings related to group differences in intelligence. Howard Gardner, for example, who is known for his theory of multiple intelligences (Gardner, 1983), writes: "I myself do not condone investigations of racial differences in intelligence, because I think that the results of these studies are likely to be incendiary" (Gardner, 2001, p. 8). He also says that we should reject the work of hereditarian IQ theorists such as Arthur Jensen (whom he calls "the bad guys") because even if they

> turn out to be more correct scientifically than I am, life is short, and we have to make choices about how we spend our time. And that's where I think the multiple intelligences way of thinking about things will continue to be useful even if the scientific evidence doesn't support it. (Gardner, 2009, 45:11)

The leading intelligence researcher Robert Sternberg (2005) argues that good science is characterized by "taste in the selection of problems to solve" (p. 295), and that it is in bad taste to investigate the genetic basis of race differences. Sam Harris—a neuroscientist famous for aggressively promoting a "reason and evidence"-based worldview—recently interviewed *Bell Curve* (Herrnstein & Murray, 1994) coauthor Charles Murray. Harris accepted all of Murray's claims about hereditarianism. Yet in the introduction to the interview he said:

> I do remain skeptical about the wisdom of looking for cross-cultural or interracial differences in things like intelligence. I'm not sure what it gets you apart from a lot of pain. So many of the topics I discussed in the podcast with Murray are not topics I would ordinarily think about, or recommend that you think about. (Harris, 2017)

Although he does not argue that the information should be suppressed, he does believe that facts about group differences conduce to "pain" and we might be better off not thinking about them.

A degree of censorship is already in operation when it comes to findings supporting hereditarianism about group differences. Mainstream media coverage of the race-and-IQ controversy almost always falsely claims that there is a consensus among the relevant experts that hereditarianism has been refuted. In fact, anonymous surveys reveal that most experts on intelligence believe that there is a genetic component to race differences (Rindermann, Becker, & Coyle, 2016, 2020, Figure 3; Snyderman & Rothman, 1987, 1988). Research supporting both hereditarian and environmentalist explanations of race differences is routinely published in major psychology journals, particularly in psychometrics journals like *Intelligence* and *Personality and Individual Differences*. But work supporting hereditarianism can be much more difficult to publish and disseminate, and research testing the possible genetic basis of race differences is rarely funded. As James Flynn (after whom the "Flynn effect" is named) notes: "If universities have their way, the necessary research [on race and intelligence] will never be done. They fund the most mundane research projects, but never seem to have funds to test for genetic differences" (Flynn, 2012, p. 36). Flynn (2018) says that "scholars at one of America's most distinguished universities...admitted [to him] that they had never approved a research grant that might clarify whether black[s] and white[s] had equivalent genes for IQ" (p. 128). When he suggested some ways to test if there is a genetic component to the black–white IQ gap, "they evaded the issue" (Flynn, personal communication). And because the state of the science is often misrepresented to the public (hence the common belief that there is no evidence at all for hereditarianism), findings are subject to de facto censorship. In some academic fields such as sociology and history, virtually all mainstream scholars refuse to consider the implications of group differences in intelligence for the problems they address. The questions before us are: Should research on group differences be conducted? Should findings related to differences be publicly disseminated? And, finally, what role should philosophers—who often act as intermediaries between science and the humanities, and between science and society at large—play in supporting or opposing work on the genetics of group differences?

Arguments in favor of restricting scientific research generally, or research on group differences specifically, are almost always utilitarian: They contend that free inquiry can, in some cases, be expected to have an overall negative effect on human well-being (e.g., Block & Dworkin, 1974; Chomsky, 1976; Dummett, 1981; Kourany, 2016). I argue that utilitarian arguments in favor of restricting research on group differences in intelligence have generally given short shrift to the potential and *actual serious harms that have resulted from the current practice of stigmatizing and dismissing controversial work*, and they have also ignored the potential benefits that might follow from openly confronting evidence about group differences.

I suggest that there are also at least two compelling *non*utilitarian reasons to allow free inquiry (compelling, that is, to those who are not utilitarians). First, truth is (to some degree) valuable independent of its felicific consequences. Proscribing the study of group differences undermines the integrity of huge swaths of work in biology, psychology, economics, sociology, history, and moral and political philosophy. If we accept that truth is intrinsically valuable, this is a cause for concern. Second, if not all groups have identical distributions of potential, then it is unjust to assume that some people must be blamed for average differences in performance among groups.

4. Utilitarian reasons to favor free inquiry

In 2001, virologists at the State University of New York at Stony Brook created a synthetic version of the poliovirus using strands of DNA purchased by mail order. They obtained the RNA sequence of the virus from the Internet, where it is freely available for all to see. Other researchers have published what are essentially instructions for recreating a smorgasbord of viruses that could potentially kill tens of millions of people were they to be released in the modern world (Kourany, 2016, pp. 784–785). In response, prominent medical authorities have called for restrictions on creating deadly viruses or publishing information that could help terrorists to do so (National Research Council, 2004).

Kourany (2016) asks: "Is there any reason new policy constraints like these for the life sciences should not be put into effect for the social sciences—new policy constraints that include research guidelines for weighing societal harms of research against societal benefits...?" (p. 786). As to the suggestion that the potential harm from deadly viruses is greater than the harm from research on intelligence differences, she urges us not to minimize the latter. Intelligence research

> has already been shown to cause significant harm...to lots more people [than a virus]—to all the people whose self-esteem, self-efficacy, ambitions, and successes are lessened as a result of direct or indirect exposure to the research or aspects of the research...; to all the people whose self-esteem, self-efficacy, ambitions, and successes are lessened as a result of the treatment they receive from others who have been directly or indirectly exposed to the research or aspects of the research; and so on—in short, to all or most women and, in the United States at least, to most minority men and many men of color in other parts of the world....Cognitive group differences

> research, then, arguably does pose harms to society near—perhaps even exceeding—the harms posed by the recent synthetic genomics research. (p. 787)

The reality of all these harms cannot, however, be taken for granted. As for the situation in the U.S., it is a well-established finding that the average levels of self-esteem in ethnic groups has, contra common wisdom, a *negative* relationship with average IQ. At least since the 1980s, Blacks have scored significantly higher than Whites, who in turn have scored higher than Asians, on measures of self-esteem (Bachman, O'Malley, Freedman-Doan, Trzesniewski, & Donnellan, 2011; Twenge & Crocker, 2002). Is the success of members of some ethnic groups impaired by hearing about research on intelligence differences? That depends on the strength of the effect of "stereotype threat." According to the theory of stereotype threat, telling people that their group has a low mean score on a cognitive test (i.e., priming them with a negative stereotype) impairs their performance (Steele & Aronson, 1995). In recent years, stereotype threat has run into trouble.

The "replication crisis" in psychology has revealed that many widely touted discoveries—particularly in social psychology and, within social psychology, *specifically* in regard to priming effects—may be completely false (Kahneman, 2012; Open Science Collaboration, 2015; Yong, 2012). Some psychologists have suggested that stereotype threat could end up being another casualty of the crisis. Ganley et al. (2013) failed to replicate the effect in three large studies of female mathematics performance. Using statistical techniques to test for publication bias, Flore and Wicherts (2015) found evidence that publication bias may "seriously distort the literature on the effects of stereotype threat" (p. 25) on math performance in girls. In a pre-registered study of more than two thousand Dutch high school students, Flore, Mulder, and Wicherts (2018) failed to find any evidence that gender stereotype threat affects math performance. There has not yet been an authoritative replication attempt of studies on the effect of priming with *racial* stereotypes. As of now, we certainly cannot say that stereotype threat "has already been shown to cause significant harm" (Kourany, 2016, p. 787). If stereotype threat is not real, then conducting research on group differences in IQ will not harm lower-scoring groups via a nonexistent effect. Even if it does turn out to be real, it is not clear that it is responsible for "significant harm," since the black–white IQ gap remains unchanged even when stereotype threat is eliminated. Stereotype threat is alleged only to *increase* the gap under controlled test-taking conditions (Sackett, Hardison, & Cullen, 2004; Steele & Aronson, 2004, p. 48). If stereotype threat turns out to be both real and to cause some devastating harm, that could raise ethical problems for intelligence research—but even its reality is far from being established.

There is also an important sense in which restrictions on publishing instructions to create deadly viruses and restrictions on intelligence research are not analogous. In the former case, the proposal is to *conceal* knowledge while being open about our motives: we should not (according to the proposals) publish instructions about how to make deadly viruses because we do not want terrorists to use them to make biological weapons. In the latter case, however, the proposal is to, if necessary, *misrepresent* the state of knowledge while being dishonest about our motives—to assert that there are no innate differences between groups while pretending that this is an established discovery of science. Given the disanalogy, we cannot so easily make an inference from the acceptability of limits on publishing work in virology to the acceptability of censoring or prohibiting work on intelligence, *even if* findings concerning intelligence differences could be shown to produce some sort of harm.

It is also easy to overlook the harms that have been caused by uncritical commitment to environmentalism. Kourany (2016) comments:

> Finding out that blacks have lower IQ scores than whites...could be the beginning of educational and training programs to work with the strengths and work on the weaknesses of every group to help make them the very best they can be, and even to use the special talents of each group to help the others. Finding these things out could be the beginning of innovative programs that support rather than undermine the right to equality. That this does not happen, or seldom happens, is a function of the...racism of society.... (pp. 783–784)

But the reason that these programs, which Kourany rightly says ought to exist, have never been created is not because of racism but *because of the taboo on talking about genetic differences among policy makers*. No mainstream politician can acknowledge that there are differences that might call for the creation of a program to "work with the strengths and work on the weaknesses of every [ethnic] group to help make them the very best they can be." It is *hereditarians* who have advocated these programs and *environmentalists* who have resisted them. The abstract to Arthur Jensen's 1969 paper—the first major modern defence of hereditarianism—says (written in the third person):

> Jensen examines other mental abilities [besides *g*] that might be capitalized on in an educational program, discussing recent findings on diverse patterns of mental abilities between ethnic groups and his own studies of associative learning abilities that are independent of social class. He concludes that educational attempts to boost IQ have been misdirected and that the educational process should focus on teaching much more specific skills. He argues that this will be accomplished most effectively if educational methods are developed which are based on other mental abilities besides IQ. (Jensen, 1969, p. 2)

Contemporary hereditarians have also called for tailored training programs (e.g., Gottfredson, 2005a; 2005b, p. 318; Lubinski & Humphreys, 1997). Accumulating evidence suggests that education can indeed be effective in raising specific cognitive skills (Ritchie, Bates, & Deary, 2015).

Decades ago, Jensen (1969) was vilified for predicting that Head Start—a program based on the environmentalist assumption that early intervention can permanently raise intelligence and academic performance—would not have its intended effects. In *The Bell Curve*, hereditarians[4] Herrnstein and Murray (1994) marshaled compelling evidence that Head Start was not working. In 2012, a congress-mandated report by the U.S. Department of Health and Human Services (HHS) stated that the effects of Head Start disappear by third grade (Puma et al., 2012, p. xxi)—that is, Head Start was not working. (Of course, the report did not endorse hereditarianism or recommend that Head Start be discontinued. It suggested that the program may confer some as-yet-undetectable benefits that last many years after participation.[5]) Nevertheless, in the years immediately following HHS's report, the government increased funding for Head Start by hundreds of millions of dollars (Enriquez, 2016). If we had followed Jensen's recommendation in 1969 to devote money to programs that were tailored to the strengths of different groups rather than to Head Start, around two hundred billion dollars would have gone to improving lives instead of accomplishing nothing that can be detected. This is one of many examples of how basing social policies on questionable scientific premises can have enormous opportunity costs.

4.1. Unknown dangers

It is impossible to know exactly what the consequences would be if race differences were proven to have a substantial genetic component and this finding became widely accepted. We ought to try to anticipate the potential harms and—keeping in mind the poor track record of even the most talented prognosticators of cultural trends—determine how seriously we should take them.

---

[4] Herrnstein and Murray (1994) conclude that "it seems highly likely...that both genes and environment have something to do with racial differences" (p. 311). They do not give an estimate as to the relative importance of one or the other.

[5] Some scholars claim that early intervention programs, such as the Abecedarian Project, have produced lasting benefits. As noted, the evidence for these claims is often questionable (see Baumeister & Bacharach, 2000).

In the liberal West, many of our institutions, laws, and moral values are predicated on the assumption that race differences are either nonexistent or environmentally caused. If the hereditarian theory of race differences became widely accepted, it seems inevitable that *there would be significant cultural changes*, even if we cannot predict what those changes would be with any precision. Just about any radical change—even if it ultimately benefits the vast majority of people—will have *some* negative consequences. But the fact that change always has negative consequences is not *by itself* a compelling reason to preserve the status quo. Every instance of progress has negative byproducts: alternative energy puts coal miners out of work, teaching evolution causes distress to creationists, and so on. Many examples throughout history show that attempts to block scientific progress, for whatever reason, are generally both futile and ultimately harmful. Is research on race differences a special case? Would finding a genetic explanation for race differences pose a uniquely serious threat?

A common fear is that, if race differences were proven to have a genetic basis, this would cause people to turn to Nazism. Indeed, the study of race differences is often explicitly equated with Nazism. This fear seems to be based on a historical misunderstanding. Nazi ideology was not based on scientific discoveries. The Nazis were flagrant pseudoscientists whose research in biology and psychology was permeated with ideology. Contrary to a popular myth, both the Nazis and their ideological predecessors (such as Joseph Arthur Comte de Gobineau and Houston Stewart Chamberlain) rejected Darwinism. (Richards, 2013, magisterially documents the lack of a connection between Darwinism and Nazism.) Most important for the present discussion, Nazi scientists rejected mainstream intelligence testing and the concept of IQ (i.e., general intelligence, or the *g* factor derived by factor analysis), preferring a mystical view of intelligence tied to race. Erich Jaensch, an influential Nazi psychologist at the University of Marburg, claimed that IQ tests advanced the "supremacy of Bourgeoisie spirit" and would be a tool "of Jewry [to] fortify its hegemony" (Rindermann, 2018, p. 61). Jaensch's student Friedrich Becker advocated "intelligence measurement according to a national and typological point of view," and called for what was essentially a version of multiple-intelligences theory that accounted for "realism," "conscientiousness," "the character value of intelligence," and "practical intelligence" (Rindermann, 2018, p. 61).

The fact that the Nazis rejected Darwinism and the concept of IQ does not necessarily mean that the study of race differences could never inspire a dangerous political movement. It does mean, however, that we cannot point to Nazi Germany as an illustration of what is liable to happen when people adopt a biologically informed view of the evolution of intelligence. In Nazi Germany, "science" followed from ideology, not the other way around. Some contemporary neo-Nazis claim to find support for their Nazi-inspired views in Darwinism and IQ research, but their ideology tends to dictate their interpretation of the science—which is usually grossly uninformed.

There is a legitimate fear that findings on race differences could be widely misunderstood in ways that would promote racial discrimination, hatred, and conflict. Our tendency for "folk essentialism," which manifests in early childhood (Gelman, 2003), is a possible source of danger. People tend to explain why each entity has the properties it does by appealing to invisible essences shared by all members of a particular category that generate those properties (Gelman, 2003). Thus, children believe that rocks, men, women, and trees belong to categories whose members possess rock-ness, maleness, etc. The fear is that the discovery of race differences could reinforce our tendency to see members of different races as possessing different essences, which would in turn foster stereotypical thinking and discrimination.

There is reason to think that this dismal outcome is not inevitable. First, cross-cultural studies have found that "essentialist intuitions [vary] enormously across cultures and education levels" (Olivola & Machery, 2014, p. 499), and are thus highly plastic. Some evidence suggests, for example, that in cultures that recognize transsexual women as a third gender people do not essentialize males (Olivola & Machery, 2014). Perhaps we should not be too pessimistic about people's capacity to learn the error of essentializing race. Second, outside of certain university departments, virtually everyone believes that there are physical and psychological differences (on average) between men and women, and most people essentialize these differences. The fact that most people in the West essentialize the differences between men and women did not stop men from supporting women's rights, and did not prevent societies from making great progress toward that end. Even if people continued to (erroneously) essentialize races, this need not create an insuperable barrier to racial harmony.

We should also keep in mind that we have to weigh the danger of open inquiry *against the alternative*, which is opposing open inquiry. If there is a legitimate fear that disseminating findings about race differences could lead people to essentialize race and then lead to racial conflict, we have to consider the possibility that *concealing* these findings could lead to a similar or perhaps worse result. As long as people believe that race differences have a purely environmental cause, differences will, in practice, most likely be attributed to racism or institutional racism. Denying the possible genetic cause of race differences will not stop people from being focused on race. If people believe that members of certain races are victimized or benefited by racism, this could also foster essentialist thinking. It may be *more* likely to lead to racial conflict, since, as discussed below, some people will invariably be *blamed*, and whole races are likely to be held collectively accountable.

5. Nonutilitarian considerations in favor of free inquiry

5.1. The intrinsic value of truth

From a utilitarian perspective, there is nothing inherently wrong with falsehood. If not "productive of any material effects, [it] can never, upon the principle of utility, constitute any offense at all" (Bentham, 1823/1876, § 16.24). However, many utilitarians take it to be an *empirical* fact that the felicific consequences of falsehood tend to be negative, and they oppose it on those grounds. Singer (1996), for example, advocates freedom of inquiry as far as *what* can be investigated, and specifically opposes restrictions on intelligence research.[6] (Naturally, he opposes *methods* of inquiry that would harm human or animal subjects without being justified on utilitarian grounds.) Bentham (1823/1876) says that "combined with other circumstances, there is scarce any sort of pernicious effect which [falsehood] may not be instrumental in producing" (§ 16.24). When it comes to extolling truth, even philosophers who are not utilitarians often emphasize its instrumental usefulness rather than its intrinsic value. Clifford (1877/1999), citing practical dangers, famously suggests that it is immoral to believe anything without sufficient evidence. He reasons that even if no harm follows from a particular false belief, we must guard against cultivating a habit of credulousness, which would eventually lead to our destruction. When Frankfurt (2006) considers why truth is important, the first thing that comes to his mind is that "truth often possesses very considerable practical utility. Any society that manages to be even minimally functional must have...a robust appreciation of the endlessly protean utility of truth" (p. 15). He expounds:

---

[6] Although it is not entirely clear what his views are, Singer (2007) appears to believe that, given the state of the science (at least in 2007), advocating hereditarianism about group differences "raise[s] the suspicion of racism" because hereditarianism "lacks a solid scientific foundation." In this passage Singer was responding to James Watson, co-discoverer of the structure of DNA, who wrote that "the overwhelming desire of society today is to assume that equal powers of reason are a universal heritage of humanity. It may well be. But simply wanting this to be the case is not enough. This is not science. To question this is not to give in to racism" (Watson, 2007). Singer (2007) agrees that "questioning this assumption is not, in itself, racist."

> Our success or failure in whatever we undertake, and therefore in life altogether, depends on whether we are guided by truth or whether we proceed in ignorance or on the basis of falsehood. [To] the extent that we recognize what dealing effectively with the problems of life entails, we cannot help loving truth. (pp. 35–36, 48)

All the reasons Frankfurt gives for why we should be concerned with truth connect directly or indirectly to its usefulness in helping us to satisfy our desires or otherwise achieve our goals.

There is also a long-standing tradition that truth has some intrinsic value, and that comprehending the truth and acting in conformity with it are worthwhile goals even if, in some cases, "he who increases knowledge increases sorrow" (Ecclesiastes 1:18). When Democritus said that he would "rather discover one causal law than be King of Persia" (Pearl, 2000, p. 41). he was probably expressing his deep commitment to truth rather than claiming that understanding leads to the greatest possible pleasure. Nozick (1974) took it for granted that most people (under normal circumstances) would reject a life of pleasure in an "experience machine" at the cost of being unaware that their perceived reality was a simulation and thus not really *true*.[7]

People do have different intuitions and views about whether truth is intrinsically valuable. But it seems likely that many professional philosophers and scientists are driven by a desire to attain some degree of true understanding of the world, ourselves, and our place in it. For these scholars, having wrong views is a tragedy even if they never suffer pain because of their mistake.

If there is meaningful, evolved, genetically based psychological diversity among geographic human populations, this has far-reaching implications for several areas of science (Winegard, Winegard, & Boutwell, 2017) as well as philosophy. We cannot understand the nature and evolution of intelligence unless we can explain why selection favored different levels of general intelligence, or specific intelligences, under different conditions. Since IQ is the best predictor of many individual and group outcomes that are studied in social science (Deary, 2012; Gottfredson, 2004; Haier, 2017; Jones, 2016; Rindermann, 2018; Warne, 2020), we will be in the dark about many sociological phenomena if we refuse to confront intelligence differences. Most theories in moral and political philosophy acknowledge the possibility of differences among individuals, but do not address the possibility of group differences. If holding ethical views that are consistent, justified, and applicable to the real world is valuable to us, then we must consider the implications of human diversity.

Those who believe the pursuit of truth to be valuable for its own sake recognize other, competing values. No sane person would pursue some interesting truth at the cost of destroying society or instigating genocide—the results feared by some opponents of research on race differences. The point is only that the intrinsic value of truth should at least be *part* of our moral calculation.

5.2. Justice

Linda Gottfredson (2005b) observes that "currently, racial parity in outcomes is often treated as the ultimate standard for fairness and lack of parity as a measure of White racism" (p. 317). Those who deny that there is evidence in favor of hereditarianism are forced to conclude that phenotypic differences between groups "must be artificial, manmade, manufactured. *Someone* must be at fault"

---

[7] "We learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it" (Nozick, 1974, p. 44).

(p. 318). But if hereditarianism is true, then it may be that no one should necessarily be blamed for different average outcomes among groups. There is no theory of justice that says it is right to falsely blame a group of people for wrongs they did not commit because confronting the exculpatory evidence causes us discomfort.

Some environmentalists may blame "structural racism" for group differences in IQ.[8] Structural racism refers to the lingering effects of past racism that are embodied in our institutions, and which have pernicious effects regardless of whether or not living people are personally racist. Although the theory that IQ gaps are due to structural racism does not imply that living people are racist, it still implies that living people should be blamed in some ways for unequal outcomes. If the theory is correct, people who benefit from structural racism have a moral obligation to oppose it and counteract its effects. So long as racial gaps remain, people will be blamed for failing to extirpate structural racism.

6. Conclusion: The responsibility of philosophers

The strategy—advocated by some influential scholars—of stigmatizing, suppressing, or downplaying evidence in favor of hereditarianism about group differences has been tried and has not worked. Research on this topic has been done and the results are widely available. Major psychology journals continue to publish work that deals openly with group differences (though researchers still debate the relative contribution of genes and environment, and the question has not been settled definitively). Any measures that would be effective in preventing further work, such as those advocated by Kourany (2016), would have to be so severe that they would only attract even more attention to the findings they aimed to suppress. Science will carry on, and these questions will be answered. We should prepare *in advance* for the possibility that the genes underlying intelligence differences will not be distributed identically among ethnic groups. Failure to do this will only create a vacuum for "cranks rather than scientists" to opine on the nature and consequences of group differences (Anomaly, 2017, p. 293). Reich (2018) warns that if scientists "willfully abstain from laying out a rational framework for discussing human differences, [they] will leave a vacuum that will be filled by pseudoscience" (p. 258).

I have argued that the usual utilitarian reasons given for restricting intelligence research are not convincing and, in fact, there are strong reasons, both utilitarian and nonutilitarian, to favor free inquiry. For philosophers specifically, there is an additional consideration. For decades, the contribution of philosophers to this debate has consisted mostly in providing alternative explanations for evidence seeming to support hereditarianism about race differences (see Sesardic, 2000, 2005), and advocating various kinds of restriction and censorship (see Cofnas, 2016). This may be because hereditarianism is controversial, and philosophers are strongly disincentivized from pursuing lines of argument that lead to truly controversial conclusions. Testifying to how serious this problem is, Jeff McMahan, Francesca Minerva, and Peter Singer recently founded the *Journal of Controversial Ideas*, which will allow scholars to publish pseudonymously. Singer (2017) commented that "it's unfortunate that such a journal should ever be considered necessary to enable controversial ideas to be published, but perhaps we have got to the point where it is." It is not clear what kind of controversial ideas Singer had in mind, and the journal has not yet released its first issue. But it is hard to find a more controversial idea than hereditarianism about race differences in intelligence.

---

[8] It is not clear why structural racism would lead to higher mean IQs for East Asians and Jews. But whether the "structural racism" explanation for group differences is plausible is not the point at issue here.

There is a danger for the philosophical community in putting our credibility on the line over the claim that race differences are entirely environmental. If work on genetics and neuroscience within the next decade produces convincing evidence that differences in measured intelligence among groups have a significant genetic component, there will be no way to conceal this information. The hereditarian explanation will have to be accepted, and people will know that philosophers were on the wrong side of the issue both scientifically and morally: scientifically, because we are supposed to be careful, disinterested commentators on scientific controversies, not activists supporting only the politically popular side; morally, because we did not help lay the groundwork for responding in a moral way to these facts that we should have known might be coming.

Acknowledgments: Thanks to Timothy Bates, Rafael De Clercq, Hrishikesh Joshi, Neven Sesardić, Omri Tal, and two anonymous reviewers for *Philosophical Psychology* for helpful comments on earlier drafts of this paper.

Open Access: © 2019 The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry, 24,* 819–827.

Anomaly, J. (2017). Race research and the ethics of belief. *Journal of Bioethical Inquiry, 14,* 287–297.

Bachman, J. G., O'Malley, P. M., Freedman-Doan, P., Trzesniewski, K. H., & Donnellan, M. B. (2011). Adolescent self-esteem: Differences by race/ethnicity, gender, and age. *Self and Identity, 10,* 445–473.

Baumeister, A. A., & Bacharach, V. R. (2000). Early generic educational intervention has no enduring effect on intelligence and does not prevent mental retardation: The infant health and development program. *Intelligence, 28,* 161–192.

Bentham, J. (1823/1876). *An introduction to the principles of morals and legislation* (New ed.). Oxford: Clarendon Press.

Block, N. J., & Dworkin, G. (1974). IQ, heritability and inequality, part 2. *Philosophy & Public Affairs, 4,* 40–99.

Ceci, S. J., & Williams, W. M. (2009). Should scientists study race and IQ? YES: The scientific truth must be pursued. *Nature, 457,* 788–789.

Chomsky, N. (1967, February 23). The responsibility of intellectuals. *The New York Review of Books.* Retrieved from http://www.nybooks.com/articles/1967/02/23/a-special-supplement-the-responsibility-of-intelle/

Chomsky, N. (1976). The fallacy of Richard Herrnstein's IQ. In N. J. Block & G. Dworkin (Eds.), *The IQ controversy: Critical readings* (pp. 285–298). New York: Pantheon Books.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures.* Cambridge, MA: MIT Press.

Clifford, W. K. (1877/1999). The ethics of belief. In *The ethics of belief and other essays* (pp. 70–96). New York: Prometheus Books.

Cofnas, N. (2016). Science is not always "self-correcting": Fact–value conflation and the study of intelligence. *Foundations of Science, 21,* 477–492.

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D.,...Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications, 9,* 2098.

Dawkins, R. (2006). *The god delusion.* New York: Houghton Mifflin.

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63,* 453–482.

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17,* 913–920.

Dummett, M. (1981). Ought research to be unrestricted? *Grazer Philosophische Studien, 12/13,* 281–298.

Edwards, A. W. F. (2003). Human genetic diversity: Lewontin's fallacy. *BioEssays, 25,* 798–801.

Enriquez, B. (2016). FY 2016 Head Start funding increase. Retrieved from https://eclkc.ohs.acf.hhs.gov/sites/default/files/docs/policy-pi/2016-08/acf-pi-hs-16-03_0.pdf

Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: A registered report. *Comprehensive Results in Social Psychology, 3,* 140–174.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53,* 25–44.

Flynn, J. R. (1980). *Race, IQ and Jensen.* London: Routledge & Kegan Paul.

Flynn, J. R. (2012). *Are we getting smarter? Rising IQ in the twenty-first century.* New York: Cambridge University Press.

Flynn, J. R. (2018). Academic freedom and race: You ought not to believe what you think may be true. *Journal of Criminal Justice, 59,* 127–131.

Frankfurt, H. G. (2006). *On truth.* New York: Knopf.

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence, 51,* 79–97.

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effect on girls' mathematics performance. *Developmental Psychology, 49,* 1886–1897.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gardner, H. (2001). The ethical responsibilities of professionals. Retrieved from http://thegoodproject.org/wp-content/uploads/2012/09/GoodWork2.pdf

Gardner, H. (2009). Intelligence: It's not just IQ. Rockefeller University, New York. Retrieved from http://www.youtube.com/watch?v=cP4CBpLNEyE&feature=related

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.

Gottfredson, L. S. (2004). Life, death, and intelligence. *Journal of Cognitive Education and Psychology, 4,* 23–46.

Gottfredson, L. S. (2005a). Suppressing intelligence research: Hurting those we intend to help. In R. H. Wright & N. A. Cummings (Eds.), *Destructive trends in mental health: The well-intentioned path to harm* (pp. 155–186). New York: Routledge.

Gottfredson, L. S. (2005b). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law, 11,* 311–319.

Gottfredson, L. S. (2010). Lessons in academic freedom as lived experience. *Personality and Individual Differences, 49,* 272–280.

Haier, R. J. (2017). *The neuroscience of intelligence.* New York: Cambridge University Press.

Harris, S. (2017). Waking up with Sam Harris #73 - Foridden knowledge (with Charles Murray). Retrieved from https://www.youtube.com/watch?v=Y1lEPQYQk8s

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Horgan, J. (2013, May 16). Should research on race and IQ be banned? *Scientific American.* Retrieved from https://blogs.scientificamerican.com/cross-check/should-research-on-race-and-iq-be-banned/

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1–123.

Jones, G. (2016). *Hive mind: How your nation's IQ matters so much more than your own.* Stanford, CA: Stanford University Press.

Kahneman, D. (2012). A proposal to deal with questions about priming effects. *Nature.* Retrieved from https://www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf

Kourany, J. A. (2016). Should some knowledge be forbidden? The case of cognitive differences research. *Philosophy of Science, 83,* 779–790.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965, volume 4* (pp. 91–195). London: Cambridge University Press.

Lewontin, R. C. (1970). Race and intelligence. *Bulletin of the Atomic Scientists, 26,* 2–8.

Lewontin, R. C. (1972). The apportionment of human diversity. *Evolutionary Biology, 6,* 381–398.

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly,* 45, 155–176.

Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176–193). Cambridge, UK: Cambridge University Press.

Lubinski, D., & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence, 24,* 159–201.

Murray, C. (2007). The magnitude and components of change in the black–white IQ difference from 1920 to 1991: A birth cohort analysis of the Woodcock–Johnson standardizations. *Intelligence, 35,* 305–318.

National Research Council. (2004). *Biotechnology research in an age of terrorism.* Washington, DC: National Academies Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J.,...Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Nisbett, R. E. (2017, June 17). Who is cherry-picking? *Vox.* Retrieved from https://www.vox.com/the-big-idea/2017/6/15/15797120/race-black-white-iq-response-critics

Nozick, R. (1974). *Anarchy, state, and utopia.* New York: Basic Books.

Olivola, C. Y., & Machery, E. (2014). Is psychological essentialism an inherent feature of human cognition? *Behavioral and Brain Sciences, 37,* 499.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349,* aac4716.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry, 20,* 98–108.

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence, 53,* 202–210.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F.,...Downer, J. (2012). *Third grade follow-up to the Head Start impact study final report.* Washington, DC: U.S. Department of Health and Human Services. Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/head_start_report.pdf

Reich, D. (2018). *Who we are and how we got here: Ancient DNA and the new science of the human past.* New York: Pantheon Books.

Richards, R. J. (2013). Was Hitler a Darwinian? In *Was Hitler a Darwinian? Disputed questions in the history of evolutionary theory* (pp. 192–242). Chicago: University of Chicago Press.

Rindermann, H. (2018). *Cognitive capitalism: Human capital and the wellbeing of nations.* Cambridge, UK: Cambridge University Press.

Rindermann, H., Becker, D., & Coyle, T. R. (2016). Survey of expert opinion on intelligence: Causes of international differences in cognitive ability tests. *Frontiers in Psychology, 7,* 399.

Rindermann, H., Becker, D., & Coyle, T. R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence, 78,* 101408.

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology, 51,* 573–582.

Rose, S. (2009). Should scientists study race and IQ? NO: Science and society do not benefit. *Nature, 457,* 786–788.

Rosenberg, N. A., Edge, M. D., Pritchard, J. K., & Feldman, M. W. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and Public Health, 2019,* 26–34.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A.,...Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50,* 912–919.

Scarr, S. (1998). On Arthur Jensen's integrity. *Intelligence, 26,* 227–232.

Sesardic, N. (1992). Science and politics: Dangerous liaisons. *Journal for General Philosophy of Science, 23,* 129–151.

Sesardic, N. (2000). Philosophy of science that ignores science: Race, IQ and heritability. *Philosophy of Science, 67,* 580–602.

Sesardic, N. (2005). *Making sense of heritability.* Cambridge, UK: Cambridge University Press.

Singer, P. (1996). Ethics and the limits of scientific freedom. *The Monist, 79,* 218–229.

Singer, P. (2007, November 1). Should we talk about race and intelligence? *Project Syndicate.* Retrieved from https://www.project-syndicate.org/commentary/should-we-talk-about-race-and-intelligence

Singer, P. (2017). What is it like to be a philosopher? Retrieved from http://www.whatisitliketobeaphilosopher.com/#/peter-singer/

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E.,...Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics, 49,* 1107–1112.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42,* 137–144.

Snyderman, M., & Rothman, S. (1988). *The IQ controversy, the media and public policy.* New Brunswick, NJ: Transaction.

Steele, C. M., & Aronson, J. A. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797–811.

Steele, C. M., & Aronson, J. A. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist, 59,* 47–48.

Sternberg, R. J. (2005). There are no public-policy implications: A reply to Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11,* 295–301.

Tal, O. (2012). The cumulative effect of genetic markers on classification performance: Insights from simple models. *Journal of Theoretical Biology, 293,* 206–218.

Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: Meta-analyses comparing Whites, Blacks, Hispanics, Asians, and American Indians and comment on Gray-Little and Hafdahl (2000). *Psychological Bulletin, 128,* 371–408.

Urbach, P. (1974). Progress and degeneration in the 'IQ debate' (I). *The British Journal for the Philosophy of Science, 25,* 99–135.

Waldman, I. D., Weinberg, R. A., & Scarr, S. (1994). Racial-group differences in IQ in the Minnesota Transracial Adoption Study: A reply to Levin and Lynn. *Intelligence, 19,* 29–44.

Warne, R. T. (2020). *In the know: Debunking 35 myths about human intelligence.* New York: Cambridge University Press.

Watson, J. (2007, October 19). To question genetic intelligence is not racism. *The Independent.* Retrieved from https://www.independent.co.uk/voices/commentators/james-watson-to-question-genetic-intelligence-is-not-racism-5328720.html

Winegard, B., Winegard, B., & Boutwell, B. (2017). Human biological and psychological diversity. *Evolutionary Psychological Science, 3,* 159–180.

Yong, E. (2012). Nobel laureate challenges psychologists to clean up their act. *Nature.* Retrieved from https://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535

CHAPTER 2

HOW STIFLING DEBATE AROUND RACE, GENES AND IQ CAN DO HARM[1]

NOAH CARL

Abstract

It is often asserted that, when it comes to taboo topics like race, genes and IQ, scholars should be held to higher evidentiary standards or even censored entirely because of the harm that might result if their findings became widely known. There is held to be an asymmetry whereby the societal costs of discussing certain topics inevitably outweigh any benefits from doing so. This paper argues that no such asymmetry has been empirically demonstrated, and that stifling debate around taboo topics can itself do active harm. To the extent that the paper's argument has force, it cannot simply be taken for granted that, when in doubt, stifling debate around taboo topics is the ethical thing to do.

Keywords*:* race, genes, IQ, ethics, free speech

> To suppress free speech is a double wrong. It violates the rights of the hearer as well as those of the speaker.
> —Frederick Douglas, 1860, *A Plea for Free Speech in Boston*

The Putative Asymmetry

It is often asserted that, when it comes to taboo topics like race, genes and IQ, scholars should be held to higher evidentiary standards or even censored entirely because of the harm that could result if their findings became widely known (see Pinker, 2002, Ch. 6; Gottfredson, 2010; Cofnas, 2016). There is held to be an asymmetry whereby the societal costs of discussing certain topics inevitably outweigh any benefits from doing so. For example, Kitcher (1985) writes,

> Everybody ought to agree that, given sufficient evidence for some hypothesis about humans, we should accept that hypothesis whatever its political implications. But the question of what counts as sufficient evidence is not independent of the political consequences. If the costs of being wrong are sufficiently high, then it is reasonable and responsible to ask for more evidence than is demanded in situations where mistakes are relatively innocuous.

Likewise, Block and Dworkin (1974) write[2],

> We are not... saying that at all times or in all places investigation of racial genotypic differences in IQ scores should stop. What we are saying is that at this time, in this country, in this political climate, individual scientists should voluntarily refrain from the investigation of genotypic racial differences in performance on IQ tests.

More recently, Gillborn (2016) writes,

---

[1] Editors' note: We have reformatted the original paper for this collection.

[2] Both this example and the example of Kitcher (1985) were taken from Cofnas (2016).

> We need to move to a position where all research on human capabilities (whether involving genetics or not) is predicated on a clear statement that any assertion of fixed and inevitable inequalities in ability/intelligence between racial/ethnic groups is, by its nature, racist.

Kourany (2016) goes further, calling for the creation of "a new National Science Advisory Board for Social Research" that would impose "tighter restrictions on race- and gender-related cognitive differences research". Her contention being that "scientists' right to freedom of research cannot be allowed to subvert other people's rights", in this case their "right to equality".

The belief that stifling debate around race, genes and IQ is necessary to prevent harm has led to scholars being mischaracterised, censored and even physically attacked. Numerous examples can be found in Pinker (2002; Ch. 6), Nyborg (2011), Winegard and Winegard (2015), and Warne *et al.* (2018). Let us consider just a few. Following the publication of his book *Sociobiology*, which applied concepts from evolutionary biology to human social behaviour, E.O. Wilson was accused of attempting to justify genocide by fellow academics, and was subjected to a campaign of harassment by irate student activists; defamatory leaflets were handed out, his lectures were invaded, and he was even doused with a jug of water (Pinker, 2002). After publishing an article in *The Harvard Educational Review* arguing that efforts to boost children's IQs and scholastic achievement were limited by their genetic endowments, Arthur Jensen was roundly denounced as a fascist, racist and elitist. In the months that followed, he was unable to open his mail, had to be escorted around the Berkeley campus by bodyguards, and was eventually forced to move house (Nyborg, 2011). When Charles Murray co-authored his book *The Bell Curve* (with Richard Herrnstein), which discussed studies of race differences in IQ, he was excoriated for supposedly trying to demonstrate that blacks were genetically inferior to whites (Winegard & Winegard 2015). Twenty-three years later, Murray was invited to give a talk at Middlebury College about an unrelated book. Soon after the talk began, he was shouted down by a jeering mob of students, so the discussion was moved to another site and broadcast via live stream. Once the discussion had finished, Murray and his interviewer Alison Stanger (a liberal Democrat who disagreed with Murray) were accosted by a group of students. As Stanger (2017) recounts,

> Most of the hatred was focused on Dr Murray, but when I took his right arm to shield him and to make sure we stayed together, the crowd turned on me. Someone pulled my hair, while others were shoving me. I feared for my life. Once we got into the car, protesters climbed on it, hitting the windows and rocking the vehicle whenever we stopped to avoid harming them. I am still wearing a neck brace, and spent a week in a dark room to recover from a concussion caused by the whiplash.

The foregoing examples illustrate the violent lengths to which some people will go to in order to stifle debate around race, genes and IQ. Why does this area of research incite such vitriolic indignation? A likely reason, as Winegard and Winegard (2015) argue, is that for a large number of academics in the West, the notion of biological sameness between groups (classes, sexes, races) has become what Tetlock (2003) calls a 'sacred value' (and see Ginges *et al.*, 2007).[3] Sacred values possess at least two important properties. First, they are incommensurable with respect to instrumental values: no amount of a sacred value can be traded off for any amount of an instrumental value. And second, proposals to accept such trade-offs are met not merely with rejection, but with moral outrage. Because arguments such as Wilson's, Jensen's and Murray's clearly threaten the sacred value of biological sameness between groups, it is not enough simply to attack the arguments; the defenders of those arguments must be hounded, and their characters impeached.

---

[3] Winegard *et al.* (2018) have dubbed this notion 'equalitarianism'.

Furthermore, there is a large body of research in psychology showing that people are quite bad at objectively appraising risk (Kahneman, 2011, Ch. 13). For example, we tend to be more afraid of snakes, spiders and large carnivores than of loaded guns, faulty electrical wires and driving without a seatbelt (Pinker, 1997, Ch. 6.) One particularly important source of error is the 'affect heuristic', whereby people judge things to have worse consequences if their mental images of those things are imbued with more negative emotional content. As Slovic *et al.* (2007) note, "activities associated with cancer are seen as riskier and more in need of regulation than activities associated with less dreaded forms of illness, injury, and death (*e.g.*, accidents)." The existence of the 'affect heuristic' should give us pause before concluding that the degree of moral outrage associated with a phenomenon constitutes a good measure of how much risk that phenomenon actually poses to society.

Although a great many areas of science (*e.g.*, the germ theory of disease, the chemistry of particulates, the psychology of manipulation) are open to misuse, there are few if any where the putative asymmetry between societal costs and scientific or other benefits is held to be as great as in the area of race, genes and IQ. Of course, the main concern among scholars who subscribe to this asymmetry is that evidence of a genetic contribution to IQ differences between human populations would be used by racists to justify oppression or exploitation of populations with lower average IQs. For example, if it were found that the difference in mean IQ between European Americans and African Americans is partly genetic, the difference would be in some sense fixed, and the worry is that racists would then have a justification for oppressing or exploiting African Americans. It goes without saying that this concern should be taken seriously; the possibility of an asymmetry between the costs and benefits of discussing race, genes and IQ is not one that should be dismissed out of hand.

However, this paper argues that no such asymmetry has been empirically demonstrated, and that stifling debate around taboo topics can itself do active harm.[4] To the extent that the paper's argument has force, it cannot simply be taken for granted that, when in doubt, stifling debate around taboo topics is the ethical thing to do. The paper makes three main claims: first, that equating particular scientific statements with racism effectively holds our morals hostage to the facts; second, that the 'blank slate' view of human nature also has pernicious moral implications; and third, that there are clear examples of where stifling debate has done material harm to both individuals and societal institutions.

Holding Our Morals Hostage to the Facts

The first way in which stifling debate around taboo topics can do harm is by holding our morals hostage to the facts. By equating particular scientific statements (*e.g.*, 'the difference in mean IQ between European Americans and African Americans may be partly genetic') with racism (*e.g.*, 'African Americans are genetically inferior to European Americans'), those seeking to stifle debate commit the moralistic fallacy of concluding that a statement cannot be true if it has unpleasant moral implications (Davis, 1978).[5] And in doing so, they make a rather perverse assumption, namely that if the relevant scientific statements *were* ever shown to be true, then the unpleasant moral implications *would* be valid. Yet as Pinker (2002; Ch. 6) notes, "We should not concede that *any* foreseeable discovery about humans could have such horrible implications… political equality is a moral stance, not an empirical hypothesis".[6]

---

[4] The present paper focuses on the harm done by stifling debate around race, genes and IQ. In a separate paper, Jeffrey and Shackelford (2017) examine "the benefits of knowing more about variance in intelligence", and put forward a "tentative case that the benefits outweigh the costs".

[5] It should be noted that not all of those who have sought to stifle debate around race, genes and IQ are guilty of this fallacy.

Indeed, there is already substantial evidence that *individuals* differ from one another with respect to IQ at least partly for genetic reasons (Plomin & Deary, 2015; Sniekers *et al*., 2017; Hill et al., 2018). But of course, this does not justify oppression or exploitation of those who have lower IQs. Likewise, there is already substantial evidence that human populations differ from one another with respect to traits like height, weight, bone density, muscle fibre distribution, lactose tolerance, thermogenic capacity and resistance to disease at least partly for genetic reasons (Epstein, 2014; Winegard *et al*., 2017). Yet, once again, this does not justify oppression or exploitation of populations who have lower means on these traits. Note that it is not being asserted that a genetic contribution to racial gaps in IQ has been conclusively demonstrated, but rather that such a finding wouldn't have a qualitatively different epistemological status from the recent finding that, say, genes associated with increased height[7] are elevated in Northern Europeans relative to Southern Europeans (Turchin *et al*., 2012; Robinson *et al*., 2015).

The point that we should not hold our morals hostage to the facts has been made over and over again by scholars interested in race, genes and IQ. For example, Wilson (1978) stated,

> Given that humankind is a biological species, it should come as no shock to find that populations are to some extent genetically diverse in the physical and mental properties underlying social behaviour. A discovery of this nature does not vitiate the ideals of Western civilisation. We are not compelled to believe in biological uniformity in order to affirm human freedom and dignity.

Similarly, Jensen (1972) stated,

> We must clearly distinguish between research on racial differences and racism. Racism implies hate or aversion and aims at denying equal rights and opportunities to persons because of their racial origin… But to fear research on genetic differences in abilities is, in a sense, to grant the racist's assumption: that if it should be established beyond reasonable doubt that there are biologically or genetically conditioned differences in mental abilities among individuals or groups, then we are justified in oppressing or exploiting those who are most limited in genetic endowment. This is, of course, a complete *non sequitur*.

Herrnstein and Murray (1994, Ch. 13) stated,

> Nothing seems more fearsome to many commentators than the possibility that ethnic and race differences have any genetic component at all. This belief is a fundamental error. Even if the differences between races were entirely genetic (which they are surely not), it should make no practical differences in how individuals deal with each other. The real danger is that the elite wisdom on ethnic differences—that such differences cannot exist—will shift to opposite and equally unjustified extremes. Open and informed discussion is the one certain way to protect society from the dangers of one extreme view or the other.

As far back as the 1960s, one of the founding fathers of the 'modern synthesis' Ernst Mayr (1963) stated,

> Equality in spite of evident non-identity is a somewhat sophisticated concept and requires a moral stature of which many individuals seem to be incapable. They rather deny human variability and equate equality with identity. Or they claim that the human species is exceptional in the organic world in that only morphological characters are controlled by genes and all other traits of the mind or character are due to "conditioning" or other

---

[6] Singer (2007) makes precisely the same point when he says, "no matter what the facts on race and intelligence turn out to be, they will not justify racial hatred, nor disrespect for people of a different race."

[7] Note also that height, just like IQ, is a *socially salient* trait. It has a robust positive association with individual income (Tyrell *et al.*, 2016), and taller men are considered more attractive than shorter men (Nettle, 2002).

> non-genetic factors… An ideology based on such obviously wrong premises can only lead to disaster. Its championship of human equality is based on a claim of identity. As soon as it is proved that the latter does not exist, the support of equality is likewise lost.

And in a recent *New York Times* op-ed, geneticist David Reich (2017) stated,

> I am worried that well-meaning people who deny the possibility of substantial biological differences among human populations are digging themselves into an indefensible position, one that will not survive the onslaught of science. I am also worried that whatever discoveries are made — and we truly have no idea yet what they will be — will be cited as "scientific proof" that racist prejudices and agendas have been correct all along, and that those well-meaning people will not understand the science well enough to push back against these claims.

As the quotes from Jensen, Mayr and Reich hint at, equating particular scientific statements with racism is not merely logically fallacious, but potentially unethical too. The reason being that it may end up encouraging precisely the behaviour that it aims to forestall. Suppose for the sake of argument that, one day in the future, evidence for a genetic contribution to psychological differences between human populations becomes so overwhelming that it cannot be reasonably denied. Note that there is nothing in science which rules this possibility out (Flynn, 2017). If between now and then, anyone who claims that genes might contribute to psychological group differences is pilloried as a 'racist', when the evidence eventually does become overwhelming, a much greater number of people are likely to take it as 'scientific proof that racism was right all along'. By contrast, if instead it is continuously asserted that "political equality is a moral stance, not an empirical hypothesis"—to quote Pinker once again—there is arguably much less danger of any discovery being taken as 'scientific proof of racism'.

The distinction between facts on the one hand, and values on the other, goes back at least to David Hume (1739; Bk. 3, Pt. 1), who demonstrated that propositions about what *ought* to be the case cannot be derived from propositions about what *is* the case (and see Moore, 1903, Ch. 1). In other words, no normative conclusion is implied by any positive proposition except in conjunction with an auxiliary normative proposition. For example, take the normative conclusion 'whites are justified in oppressing and exploiting blacks'. There is no pair of positive propositions from which this conclusion can be validly deduced. For example, the argument 'blacks have a lower mean IQ than whites; therefore whites are justified in oppressing and exploiting blacks' is obviously invalid. Indeed, it is not even a syllogism. The conclusion 'whites are justified in oppressing and exploiting blacks' can only be derived from the premise 'blacks have a lower mean IQ than whites' in conjunction with *another normative proposition*, namely that 'races with higher average IQs are justified in oppressing or exploiting races with lower average IQs'. And there is of course no scientific evidence that could be adduced in support of that proposition. To summarise, particular scientific findings (*e.g.*, that variation in IQ is partly genetic) are logically independent of particular normative conclusions (*e.g.*, that people should not be exploited).

One possible objection to the preceding argument goes as follows. The discovery of a genetic contribution to, say, race differences in crime rates would seem to provide a strong Bayesian rationale for policies like racial profiling (*e.g.*, stopping and searching black youths more often than white youths), which many people believe are unethical (see Risse & Zeckhauser, 2004; Sesardić, 2018). Insofar as this is the case, can it really be said that scientific findings are logically independent of normative conclusions? Taking 'independent' to mean 'not logically *deducible* from', yes it can. Just because something is 'rational' does not mean it is necessarily ethical. Indeed, discriminating on the basis of group differences is 'rational' regardless of whether those differences are genetic or environmental in origin, so the fact that racial profiling has already been criticised (*i.e.*, has been criticised in a context where race differences in crime rates are almost universally assumed to be environmental) illustrates that many people accept the principle that something can be unethical even when it is 'rational'. Another point worth making is that racial profiling is fundamentally different from the kinds of oppression and exploitation that were meted out by racists in the past. As Sesardić (2018) notes, the police already profile men (who are much more likely to commit crime than women), but most people don't see any problem with this. Moreover, several black scholars have pointed out that the primary beneficiaries of racial profiling are law-abiding black people who live in the neighbourhoods where black criminals operate (*e.g.*, Riley, 2015, Ch. 3; Williams, 2017).

Furthermore, some radical political philosophies such as 'luck egalitarianism' are arguably just as easy to reconcile with the findings of differential psychology and behavioural genetics as any 'elitist' political philosophy (Dworkin, 1981a,b; Cohen, 1989; Anderson, 1999; Knight, 2013).[8] For example, the philosopher Ronald Dworkin (1981a,b) drew a distinction between 'tastes and ambitions' on the one hand and 'endowments' on the other (see Lamont & Favor, 2017). An individual's tastes and ambitions comprise all the choices he makes that affect his material well-being, such as how hard to work, and whether to spend money frivolously or prudently. By contrast, his 'endowments' comprise all the handicaps or advantages he possesses due to circumstances beyond his control, such as his genes, his family's wealth or simple bad luck (*e.g.*, getting hit by a car). According to Dworkin, material inequalities that arise due to differences in 'tastes and ambitions' may be justified, but those that arise due to differences in 'endowments' are not. So even if IQ were 100% genetic (which incidentally, it is not), since individuals cannot control the genes they will inherit, there is a relatively strong case to be made that inequalities arising due to differences in IQ (whether between individuals, classes or races) should be reduced or eliminated. Of course, this is not to say that luck egalitarianism is *the* correct moral theory, but simply that the findings of differential psychology and behavioural genetics need not be inconsistent with left-wing political ideals (see Dillow, 2018).

The Pernicious Implications of the Blank Slate

The second way in which stifling debate around taboo topics can do harm is by promoting an alternative, 'blank slate' view of human nature, which in the past has proven no less pernicious in its supposed moral implications than the 'hereditarian' view we have been discussing so far. This section quotes extensively from Pinker (2002, Ch. 8), who has elucidated the argument particularly well:

> Some people have suggested to me that these grandiloquent arguments are just too fancy for the dangerous world we live in. Granted, there is evidence that people are different, but since data in the social sciences are never perfect, and since a conclusion of inequality might be used to the worst ends by bigots or Social Darwinists, shouldn't we err on the side of caution and stick with the null hypothesis that people are identical? Some believe that even if we were certain that people differ genetically, we might still want to promulgate the fiction that they are the same, because it is less open to abuse.
>
> This argument is based on the fallacy that the Blank Slate has nothing but good moral implications and a theory of human nature nothing but bad ones. In the case of human differences, as in the case of human universals, the dangers go both ways…

---

[8] As Rushton (1996) notes, "There are no necessary policies which flow from race research. The findings are compatible with a wide range of recommendations: from social segregation, through laissez-faire, to programs for the disadvantaged".

The first way in which the 'blank slate' view of human nature has been misused is as a justification for totalitarian efforts to 'remake humanity'. If there is no such thing as human nature, and individuals are not constrained by any kind of genetic endowment—so the arguments goes—those who exhibit undesirable traits can and should be 'perfected' through appropriately targeted state intervention. In the 20th century's most repressive communist dictatorships, such intervention took a number of different forms, including forced labour programs, 're-education' camps, and mass killings of the ones who proved incorrigible. As Pinker (2002, Ch. 8) notes, various communist ideologues' commitment to the 'blank slate' view of human nature is evidenced in the statements that they made:

> Lenin endorsed Nikolai Bukharin's ideal of "the manufacturing of Communist man out of the human material of the capitalist age." Lenin's admirer Maxim Gorky wrote, "The working classes are to Lenin what minerals are to the metallurgist" and, "Human raw material is immeasurably more difficult to work with than wood" (the latter while admiring a canal built by slave labor). We come across the metaphor of the blank slate in the writings of a man who may have been responsible for sixty-five million deaths:
> "A blank sheet of paper has no blotches, and so the newest and most beautiful words can be written on it, the newest and most beautiful pictures can be painted on it."
> —Mao Zedong
>
> And we find it in a saying of a political movement that killed a quarter of its countrymen:
> "Only the newborn baby is spotless."
> —Khmer Rouge slogan

The second way in which the 'blank slate' view of human nature has been misused is as a justification for persecution or genocide of 'successful' groups within society. According to this line of argument, if all groups start out identical, but some end up wealthier than others, then those who managed to accrue more wealth must have done so through nefarious means; they must have been more cunning, more avaricious or more sneaky than their rivals. Throughout the last two centuries, numerous ethnic groups and social classes have been persecuted because their success was taken as evidence of their wickedness: 'bourgeois peasants' in the Soviet Union; literate professionals in Cambodia; 'rich peasants' in Mao's China; the Indians in Uganda; the Chinese in Indonesia; the Armenians in Turkey; the Igbos in Nigeria; and the Jews in Europe, Russia and the Middle East (Pinker, 2002, Ch. 8; and see Cofnas, 2017).

Of course, the fact that the 'blank slate' view of human nature has been used to justify crimes such as genocide, forced labour and mass population displacement does not imply that the 'hereditarian' view is correct. Nor does it imply that the 'hereditarian' view hasn't also been grossly misused (most appallingly, by the Nazis). However, it does call into question the putative asymmetry whereby the societal costs of discussing genes, race and IQ are assumed to inevitably outweigh any benefits from doing so. As Pinker (2002, Ch. 8) notes, the "realization that government-sponsored mass murder can come from an anti-innatist belief system as easily as from an innatist one upends the postwar understanding that biological approaches to behavior are uniquely sinister".

Material Harm Done by Stifling Debate

Evidence presented in this section illustrates how stifling debate around taboo topics can do harm by leading scholars to ignore salient aspects of phenomena, by causing authorities to 'sweep problems under the rug', and by inciting a backlash from enraged dissidents. First of all, consider race differences in optimal medicine. Back in the 1990s, Rushton (1996) pointed out that the medical needs of people from different racial backgrounds are almost certainly not identical (and see Rushton & Jensen, 2005). In particular, he noted that

> Just as women doctors have advocated that to conceptualize women as being the same as men leads to a neglect of women's problems and their treatment (e.g., premenstrual symptoms and menopause and hormone replacement therapy), so Black doctors are increasingly becoming concerned that treating Blacks the same as Whites is to neglect Black problems

And indeed, Rushton's contention has been borne out by subsequent evidence (see Taylor & Ellis, 2002; Burroughs *et al.*, 2002; Ramamoorthy *et al.*, 2015). For example, in their 2002 review of the literature, Burroughs *et al.* concluded that

> Pharmacogenetic research in the past few decades has uncovered significant differences among racial and ethnic groups in the metabolism, clinical effectiveness, and side-effect profiles of many clinically important drugs. These differences must be taken into account in the design of cost management policies such as formulary implementation, therapeutic substitution and step-care protocols.

To deny the existence of race, or to insist that it is a wholly 'social construct', as many critics of the 'hereditarian' view of human nature do (see Wade, 2014, Ch. 5; Winegard et al., 2017; Sesardic, 2005, Ch. 4), is to commit oneself to an erroneous view of modern medicine, one which could conceivably come at the cost of people's lives.

The remaining two examples in this section do not pertain directly to the topic of race, genes and IQ. But they illustrate how stifling debate around taboo topics through 'political correctness' can do *material harm* to both individuals and societal institutions. First, it has emerged in Britain over the last few years that hundreds of vulnerable girls were systematically 'groomed' and in some cases violently gang-raped by groups of men from their local areas (*e.g.*, Jay, 2013). Such cases have been reported in: Aylesbury, Banbury, Bristol, Derby, Halifax, Keighley, Newcastle, Oxford, Peterborough, Rochdale, Telford, and Rotherham. One of the salient facts about all these cases is that the vast majority of perpetrators[9] were men of South Asian origin, whereas the victims were mostly White British girls (Rafiq & Adil, 2017). Since the cases began emerging, a number of independent government reports have been published (Jay, 2013; House of Commons, 2013; Bedford, 2015). These reports explain how various parties (such as the police, the social services and the local council) failed to intervene to stop the abuse because of misplaced concerns about 'political correctness' or for fear of being called 'racist'.[10] For example, Jay (2013), who led the independent inquiry into sexual abuse in Rotherham, notes:

> Several staff described their nervousness about identifying the ethnic origins of perpetrators for fear of being thought racist; others remembered clear direction from their managers not to do so.

Likewise, Bedford (2015), who led the independent inquiry into sexual abuse in Oxfordshire notes:

> A parent told a police station about information provided by the daughter and queried why no immediate arrests were being made. The parent says the desk officer responded by saying that such arrests could not simply be made on such information and that the Police were also under pressure not to appear institutionally racist.

---

[9] Researchers from the think tank Quilliam analysed data on grooming gang cases in the UK since 2005, and concluded that 84% of offenders were men of South Asian (mostly Pakistani) origin, despite the fact that South Asians make up no more than 7% of the British population.

[10] Similarly, in his review into electoral fraud in Britain, Sir Eric Pickles (2016) acknowledged concerns that "state institutions had turned a blind eye to such behavior because of 'politically correct' over-sensitivities about ethnicity and religion" (and see Carl, 2017).

The House of Commons report published in June of 2013 including the following among its list of recommendations for how to prevent future abuse:

> It is important that police, social workers and others be able to raise their concerns freely, without fear of being labelled racist.

The evidence from grooming gangs in Britain illustrates that throwing around unsubstantiated charges of 'racism' can create a climate of fear in which people feel too paralysed to act, and that insofar as this as this is the case, doing so should not be considered a sensible precaution, or even a mild nuisance, but a potentially unethical thing to do. Of course, it does not follow that there should be open season to make insulting or offensive remarks about ethnic minorities. The point is that there are costs as well as benefits to levelling accusations of 'racism', and that noticing patterns in the data should be clearly demarcated from intentionally denigrating whole communities.

Finally, it seems safe to assume that most of the people who are not already convinced by the arguments in this paper are unlikely to be supporters of US President Donald Trump (see Schaffner et al., 2017; Pettigrew, 2017). They would presumably, and indeed with some justification, regard his election as something that has done material harm to both individuals (*e.g.*, illegal/undocumented immigrants) and societal institutions (e.g., the Office of the Presidency). It is therefore noteworthy that opposition to 'political correctness' appears to have been a major contributor to his election (Goldberg, 2018). One Trump supporter told a journalist at *Reason* magazine (Soave, 2016), "This blind adherence to political correctness was my main issue in the recent political arena". And another told the *Washington Post* (2016), "I am a gay millennial woman and I voted for Donald Trump because I oppose the political correctness movement". Furthermore, on a poll taken prior to the presidential election, Pew Research asked Americans whether "too many people are easily offended these days over the language that others use" (Fingerhut, 2016). A full 83% of Trump supporters agreed, compared to only 39% of Clinton supporters. Using multivariate analysis, Goldberg (2018) found that opposition to 'political correctness' remained a significant predictor of support for Trump even after controlling for several measures of prejudice, as well as attitudes to immigration. And in a recent experimental study, Conway et al. (2017) observed that priming respondents with a vignette about 'political correctness' led to a significant rise in support for Trump, but no rise in support for Clinton. According to the authors, their study "provides evidence that norms that are designed to increase the overall amount of positive communication can actually backfire by increasing support for a politician who uses extremely negative language that explicitly violates the norm". To the extent that many people concerned about the interests of disadvantaged groups regard the election of Donald Trump as a negative outcome (something which seems altogether plausible), the evidence presented above suggests that stifling debate stifling debate around taboo topics through 'political correctness' may very well be *counter*-productive.

Conclusion

It is often asserted that, when it comes to taboo topics like race, genes and IQ, scholars should be held to higher evidentiary standards or even censored entirely because of the harm that might result if their findings became widely known. There is held to be an asymmetry whereby the societal costs of discussing certain topics inevitably outweigh any benefits from doing so. This paper argued that no such asymmetry has been empirically demonstrated, and that stifling debate around taboo topics can itself do active harm. To the extent that the paper's argument has force, it cannot simply be taken for granted that, when in doubt, stifling debate around taboo topics is the ethical thing to do. The argument comprised three main claims: first, that equating particular scientific statements with racism effectively holds our morals hostage to the facts; second, that the 'blank slate' view of human nature also has pernicious moral implications; and third, that there are clear examples of where stifling debate around taboo topics has done material harm to both individuals and societal institutions.

An important caveat is that the paper did not show that the societal benefits of discussing races, genes and IQ *actually do* outweigh the costs. It simply pointed out that the asymmetry assumed by those seeking to stifle debate has never been empirically demonstrated, and that there are in fact a number of ways in which stifling debate can do active harm. Consequently, the paper's overall conclusion was not that stifling debate is necessarily an *unethical* thing to do. Rather, it was a slightly weaker claim, namely that *one cannot take for granted* that stifling debate is the *ethical* thing to do. This weaker claim seems eminently reasonable, given that—at the very least—there are now a number of arguments for stifling debate around taboo topics (Kitcher, 1985; Block and Dworkin, 1974; Gillborn, 2016; Kourany, 2016), as well as a number of arguments against doing so, including the ones advanced here (and see Pinker, 2002; Flynn, 2017; Jeffrey & Shackelford, 2017).

Going forward, what steps can be taken to increase the quality of debate around taboo topics like race, genes and IQ? Most importantly, it should be reaffirmed that no normative conclusions follow from any particular scientific findings, and that holding our morals hostage to the facts is a great mistake (Pinker, 2002, Ch. 6). As noted in Section 2, evidence suggests that individuals differ from one another with respect to IQ at least partly for genetic reasons, but this does not justify oppression or exploitation of those who have lower IQs. Evidence also suggests that human populations differ from one another with respect to numerous physiological traits at least partly for genetic reasons, but this does not justify oppression or exploitation of populations who have lower means on these traits. Consequently, if it were one day conclusively demonstrated that genes contribute to psychological differences between human populations, it would not suddenly become justified to oppress or exploit the populations who scored lower on those psychological traits (Winegard *et al.*, 2017). Admitting the arguments in this paper, another step that can be taken to increase the quality of debate is to encourage persons who disagree with one another to collaborate on a piece of research that might resolve their dispute (see Duarte *et al.*, 2015).

It is worth ending with the words of philosopher Peter Singer (2008), who remarked in relation to the present debate, that "when faced with… major social problems, a preference for ignorance over knowledge is difficult to defend."

Acknowledgements

I would like to thank Jam Kraprayoon, Nathan Cofnas, Ben Winegard and an anonymous reviewer for comments that improved the manuscript.

Open Access: This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http:/ /creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Anderson, E.S. (1999). What is the Point of Equality? *Ethics, 109,* 287–337.

Bedford, A. (2015). Serious Case Review into Child Sexual Exploitation in Oxfordshire: from the experiences of Children A, B, C, D, E, and F. *Oxfordshire Safeguarding Children Board*, published online.

Block, N. J., & Dworkin, G. (1974). IQ, heritability and inequality, part 2. *Philosophy & Public Affairs, 4,* 40–99.

Burroughs, V.J., Maxey, R.W. & Levy, R. (2002). Racial and ethnic differences in response to medicines: towards individualized pharmaceutical treatment. *Journal of the National Medical Association, 94,* 1-26.

Carl, N. (2017). Ethnicity and Electoral Fraud in Britain. *Electoral Studies, 50,* 128-136.

Cofnas, N. (2016). Science is not always "self-correcting": Fact-value conflation and the study of Intelligence. *Foundations of Science, 21,* 477-492.

Cofnas, N. (2017). Judaism as a Group Evolutionary Strategy: A Critical Analysis of Kevin MacDonald's Theory. *Human Nature*, early online view.

Cohen, G.A. (1989). On the Currency of Egalitarian Justice. *Ethics, 99,* 906–44.

Conway, L.G., Repke, M.A. & Houck, S.C. (2017). Donald Trump as a Cultural Revolt Against Perceived Communication Restriction: Priming Political Correctness Norms Causes More Trump Support. *Journal of Social and Political Psychology*, 2195-3325, published online.

Davis, B.B. (1978). The moralistic fallacy. *Nature, 272*, 390.

Dillow, C. (2018). Genes & The Left. *Stumbling and Mumbling*, January 16, published online.

Duarte, J.L., Crawford, J.T., Stern, C., Haidt, J. (2015). Political diversity will improve social psychological science. *Behavioural and Brain Sciences, 38*, 1-58.

Dworkin, R. (1981a). What is Equality? Part 1: Equality of Welfare. *Philosophy and Public Affairs, 10*, 283–345.

Dworkin, R. (1981b). What is Equality? Part 2: Equality of Resources. *Philosophy and Public Affairs, 10*, 185–246.

Epstein, D. (2014). *The Sports Gene: Inside the Science of Extraordinary Athletic Performance.* London, UK: Penguin.

Fingerhut, H. (2016). In 'political correctness' debate, most Americans think too many people are easily offended. *Pew Research*, published online.

Flynn, J. (2017). Academic freedom and race: You ought not to believe what you think may be true. *Journal of Criminal Justice*, in press.

Gillborn, D. (2016). Softly, softly: Genetics, intelligence and the hidden racism of the new geneism. *Journal of Education Policy, 31*, 365-388.

Ginges, J., Atran, S., Medin, D. & Shikaki, M. (2007). Sacred bounds on rational resolution of violent political conflict. *PNAS, 104,* 7357-7360.

Goldberg, Z. (2018). Serwer Error: Misunderstanding Trump Voters. *Quillette, January 1*, published online.

Gottfredson, L.S. (2010). Lessons in academic freedom as lived experience. *Personality and Individual Differences, 49*, 272-280.

Herrnstein, R. J. & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York City, NY: Free Press.

Hill, W.D., Marioni, R.E., Maghzian, O., Ritchie, S.J., Hagenaars, S.P. *et al.* (2018). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry*, published online.

House of Commons. (2013). Child sexual exploitation and the response to localised grooming: follow-up, Second Report of Session 2013–14. *House of Commons Home Affairs Committee*, published online.

Hume, D. (1739). *A Treatise of Human Nature.* London, UK: John Noon.

Jay, A. (2013). Independent Inquiry into Child Sexual Exploitation in Rotherham. *Rotherham Metropolitan Borough Council*, published online.

Jeffrey, A.J. & Shackelford, T.K. (2017). Moral positions on publishing race differences in intelligence. *Journal of Criminal Justice*, in press.

Jensen, A. (1972). *Genetics and Education.* New York City, NY: Harper & Row Publishers.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York City, NY: Farrar, Straus and Giroux.

Kitcher, P. (1985). *Vaulting ambition: Sociobiology and the quest for human nature.* Cambridge, MA: MIT Press.

Knight, C. (2013). Luck egalitarianism. *Philosophy Compass, 8/10*, 924–934.

Kourany, J. A. (2016). Should some knowledge be forbidden? The case of cognitive differences research. *Philosophy of Science, 83,* 779 –790.

Nettle, D. (2002). Height and reproductive success in a cohort of British men. *Human Nature, 13,* 473-491.

Nyborg, H. (2011). The greatest collective scientific fraud of the 20th century: The demolition of differential psychology and eugenics. *Mankind Quarterly, 51*, 241-268.

Lamont, J. & Favor, C. (2017). Distributive Justice: Equality of Opportunity and Luck Egalitarianism. *Stanford Encyclopaedia of Philosophy*, published online.

Mayr, E. (1963). *Animal Species and Evolution.* Cambridge, MA: Harvard University Press.

Mill, J.S. (1859). *On Liberty.* London, UK: John W. Parker and Son.

Pettigrew, T.F. (2017). Social Psychological Perspectives on Trump Supporters. *Journal of Social and Political Psychology,* 2195-3325, published online.

Pickles, Sir Eric. (2016). Securing the ballot: Report of Sir Eric Pickles' review into electoral fraud. *UK Government*, published online.

Pinker, S. (1997). *How the Mind Works*. New York City, NY: W.W. Norton & Company.

Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. London, UK: Penguin Books.

Plomin, R. & Deary, I. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry, 20*, 98-108.

Rafiq, H. & Adil, M. (2017). Group Based Child Sexual Exploitation – Dissecting Grooming Gangs. *Quilliam Foundation,* published online.

Ramamoorthy, A., Pacanowski, M.A., Bull, J. & Zhang, L. (2015). Racial/ethnic differences in drug disposition and response: review of recently approved drugs. *Clinical Pharmacology & Therapeutics, 97,* 263-273.

Reich, D. (2017). How Genetics Is Changing Our Understanding of 'Race'. *New York Times*, March 23, published online.

Riley, J.L. (2015). *Please Stop Helping Us: How Liberals Make it Harder for Blacks to Succeed*. New York City, NY: Encounter Books.

Risse, M. & Zeckhauser, R. (2004). Racial profiling. *Philosophy & Public Affairs, 32*, 131-170.

Robinson, M.R., Hemani, G., Medina-Gomez, C., Mezzavilla, M. & Esko, T. *et al.* (2015). Population genetic differentiation of height and body mass index across Europe. *Nature Genetics, 47,* 1357-1362.

Rushton, J.P. (1996). Political Correctness and the Study of Racial Differences. *Journal of Social Distress and the Homeless, 5*, 213-229.

Rushton, J.P. & Jensen, A.R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy and Law, 11*, 235-294.

Schaffner, B.F., MacWilliams, M. & Nteta, T. (2017). Explaining White Polarization in the 2016 Vote for President: The Sobering Role of Racism and Sexism. *Working Paper*, published online.

Sesardić, N. (2005). *Making Sense of Heritability.* Cambridge, UK: Cambridge University Press.

Sesardić, N. (2018). Is Racial Profiling a Legitimate Strategy in the Fight against Violent Crime? *Philosophia*, 1-19.

Singer, P. (2007). Should We Talk About Race and Intelligence? *Project Syndicate,* November 1, published online.

Slovic, P., Finucane, M.L., Peters, E. & MacGregor, D.G. (2007). The affect heuristic. *European Journal of Operational Research, 177*, 1333–1352.

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R.I. et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics, 49,* 1107-1112.

Soave, R. (2016). Yes, Political Correctness Helped Elect Trump: What Skeptics Need to Know. *Reason, Hit & Run Blog*, December 19, published online.

Stranger, A. (2017). Understanding the Angry Mob at Middlebury That Gave Me a Concussion. *The New York Times*, March 13, published online.

Taylor, J.S. & Ellis, G.R. (2002). Racial differences in responses to drug treatment: implications for pharmacotherapy of heart failure. *American Journal of Cardiovascular Drugs, 2,* 389-399.

Tetlock, P.E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *TRENDS in Cognitive Sciences, 7,* 320-324.

Turchin, M.C., Chiang, C.W.K., Palmer, C.D., Sankararaman, S., Reich, D., *et al.* (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics, 44,* 1015–1019.

Tyrell, J., Jones, S.E., Beaumont, R., Astley, C.M., Lovell, R. *et al.* (2016). Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *BMJ, 352,* 1-10.

Wade, N. (2014). *A Troublesome Inheritance: Genes, Race and Human History.* London, UK: Penguin.

Warne, R.T., Astle, M.C. & Hill, J.C. (2018). What Do Undergraduates Learn About Human Intelligence? An Analysis of Introductory Psychology Textbooks. *Archives of Scientific Psychology, 6,* 32–50.

Washington Post. (2016). Why I voted for Trump. *Washington Post, Opinion*, November 23, published online.

Williams, W.E. (2017). Blacks suffer when police are undermined. *The News Virginian*, *Opinion*, February 19, published online.

Wilson, E. O. (1978). *On human nature.* Cambridge, MA: Harvard University Press.

Winegard, B., & Winegard, B. (2015). A social science without sacred values. *Working Paper,* published online.

Winegard, B., Winegard, B. & Boutwell, B. (2017). Human Biological and Psychological Diversity. *Evolutionary Psychological Science, 3,* 159-180.

Winegard, B.M., Clark, C.J. & Hasty, C. (2018). Equalitarianism: A source of liberal bias. *Working paper*, published online

CHAPTER 3

BETWEEN-GROUP MEAN DIFFERENCES IN INTELLIGENCE IN THE UNITED STATES ARE >0% GENETICALLY CAUSED: FIVE CONVERGING LINES OF EVIDENCE

RUSSELL T. WARNE

Abstract

The past 30 years of research in intelligence has produced a wealth of knowledge about the causes and consequences of differences in intelligence between individuals, and today mainstream opinion is that individual differences in intelligence are caused by both genetic and environmental influences. Much more contentious is the discussion over the cause of mean intelligence differences between racial or ethnic groups. In contrast to the general consensus that interindividual differences are both genetic and environmental in origin, some claim that mean intelligence differences between racial groups are completely environmental in origin, whereas others postulate a mix of genetic and environmental causes. In this article I discuss 5 lines of research that provide evidence that mean differences in intelligence between racial and ethnic groups are partially genetic. These lines of evidence are findings in support of Spearman's hypothesis, consistent results from tests of measurement invariance across American racial groups, the mathematical relationship that exists for between-group and within-group sources of heritability, genomic data derived from genome-wide association studies of intelligence and polygenic scores applied to diverse samples, and admixture studies. I also discuss future potential lines of evidence regarding the causes of average group differences across racial groups. However, the data are not fully conclusive, and the exact degree to which genes influence intergroup mean differences in intelligence is not known. This discussion applies only to native English speakers born in the United States and not necessarily to any other human populations.

Keywords: intelligence, IQ, group differences, behavioral genetics, race

Few constructs in the social sciences are as well understood as intelligence. One widely accepted definition states that:

> Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do. (Gottfredson, 1997a, p. 13)

The major foundational breakthroughs in the study of intelligence occurred almost simultaneously in two different countries. Spearman's development of factor analysis when analyzing educational data led him to conclude that "all branches of intellectual activity have in common one fundamental function" (Spearman, 1904, p. 284), which he called "*g*" and believed was synonymous with general intelligence. The next year Binet and Simon (1905/1916) created the first successful intelligence test, which allowed users to classify examinees by their degree of intellectual disability.

In the ensuing century, psychologists and other scientists have built on these advances to create some of the most sophisticated psychometric instruments in existence and to learn more about the nature of human intelligence (Deary, 2012). As a result, psychologists know more about intelligence—its nature, causes, and consequences—than almost any other psychological construct (Warne, Astle, & Hill, 2018). For example, research in neuroscience has strongly suggested that intelligence arises from a distributed network of well-connected brain regions, located mostly in the parietal and frontal lobes (Deary, Penke, & Johnson, 2010; Haier, 2017; Jung & Haier, 2007; Penke et al., 2012). In turn, these regions—and the brain as a whole—seem to be larger, healthier, and more organized in brighter people (Flashman et al., 1997; Genç et al., 2018; Gignac & Bates, 2017; W. D. Hill et al., 2019). Geneticists have identified specific genetic variances correlated with neurodevelopment and brain functioning that are also associated with interindividual differences in intelligence or educational attainment (e.g., Adams et al., 2016; Lee et al., 2018; Okbay et al., 2016). Likewise, intelligence differences have been shown to correlate with many important health, occupational, educational, and psychosocial outcomes (Gottfredson, 1997b; Jensen, 1998; Warne, 2016a), many of which intelligence tests were never designed to predict, such as longevity (Arden et al., 2016; Gottfredson & Deary, 2004).

Individual Differences in Intelligence

One consistent finding—first shown more than 90 years ago (Burks 1928/1973)—is the heritability of intelligence, which ranges between .20 and .80 in most studies conducted in industrialized countries (Plomin, DeFries, Knopik, & Neiderhiser, 2016). Heritability is the proportion of phenotype variance that is associated with genotype variance in a sample. Zero heritability indicates that a trait's variance is entirely environmental in origin, and a heritability value of 1.0 occurs when a trait's variance is entirely genetic. Almost every trait's heritability falls between these two extremes. For intelligence, age has been shown to be a moderator variable, with most studies showing that heritability increases with age (both longitudinally and in cross-sectional comparisons of cohorts), indicating that genetics have a strong influence on the development of intelligence in adulthood, whereas in childhood environment is more important (Bouchard, 2014; Mollon et al., 2021; Plomin & Deary, 2015; Plomin et al., 2016). These heritability estimates are derived from studies of twins, adoptees and their parents, and other family relationships, but these studies are not capable of identifying specific genetic variants that correlate with intelligence phenotypes. Genome-wide association studies (GWASs) are used for this purpose because they take advantage of data mining techniques to identify genetic variants that are correlated with a phenotype. Using GWASs, behavioral geneticists have thus far identified hundreds of portions of DNA that are correlated with intelligence phenotypes (Savage et al., 2018). More genetic variants will probably emerge from future studies with larger sample sizes, greater statistical power, and more sophisticated genome sequencing technology (Murray, 2020; Plomin & von Stumm, 2018).

Although the evidence of a genetic influence on interindividual differences in intelligence is strong, this is not to say that the environment is not important. No study has ever shown a heritability of 1.0 for intelligence, and no theorist expects such (Plomin et al., 2016). This leaves room for environmental influences, especially during childhood. Progress in identifying specific environmental influences on intelligence is not as advanced as the research on genetic influences. The lower level of certainty about environmental impacts on intelligence is not caused by a lack of trying to discover these influences. Rather, there are three main impediments. First, there is no strong theory about the nature of environment, how influence is transmitted from one generation to another, or how to measure it (Hunt, 2011). On the other hand, biology has clear answers (either empirically or theoretically) about these issues for genes.

A second difficulty is that the important environmental influences on intelligence seem to be mostly nonshared influences within families. In other words, the major environmental determinants of intelligence seem to be unique individual experiences and not any influence that family members share as they live and grow up together (Plomin et al., 2016). The idiosyncrasies of these components of the environment make these causes of intelligence differences difficult to isolate, although some have tried to do so (e.g., Asbury et al., 2016).

A final difficulty in studying environment is that many so-called environmental variables are actually partially genetically caused (Kendler & Baker, 2007). This is especially true for people raised by one or both biological parents; these people share genes with their parents, and those genes, in turn, shape the home environment via parental actions (Plomin et al., 2016). This means that many of the usual explanations of environmental influence on intelligence—such as socioeconomic status or nutrition—are confounded with genetic influence. Setting aside genetic impacts on the home environment, individuals' genes help them choose an environment that is suited to their (genetically influenced) phenotypes, a tendency that accelerates as individuals have more freedom to choose environments. This may explain why genetic influences on intelligence tend to increase in adulthood (Bouchard, 2014). Despite these genetic influences on intelligence and the fact that the exact environmental variables that are important for explaining individual differences in intelligence are not fully known, it is clear that genes are not the only influence on intelligence and that environmental characteristics do matter (Gottfredson, 1997a; Neisser et al., 1996; Warne, 2020b).

Between-Group Differences in Intelligence

Just as for individual-level differences, researchers widely recognize that racial and ethnic groups differ in their mean level of intelligence (Neisser et al., 1996; Warne et al., 2018), although there is a great deal of overlap between the distributions of intelligence in these various groups (Gottfredson, 1997a; Warne, 2020a). Average score gaps in IQ and similar measures persist in the United States, with the average IQ scores being approximately 105 for Asian Americans, 100 for White Americans, 90 for Hispanic Americans, and 85 for African Americans (see Curran & Kellogg, 2016; Dickens & Flynn, 2006; Hunt, 2011; Little, 2017; Murray, 2007, 2020, 2021 for more exact estimates aggregated from representative samples).

Although knowledge of the causes of individual differences in intelligence is increasing every year, the information on causes of mean group differences in intelligence is not as advanced. This discrepancy is at least partially due to the fact that the topic of race differences in intelligence is one of the most controversial in all of science (Cofnas, 2016; Reynolds, 2000). Indeed, because of the social sensitivity of the topic, some have argued that the very act of investigating causes of these differences is itself inherently immoral or racist (e.g., Gould, 1981; Kourany, 2016; Sternberg, 2005; Turkheimer, 2007). I reject this censorious perspective as being antithetical to science. After all, "Without a forum for controversy, controversy will not be resolved and science will not advance" (Detterman, 2006, p. iv). Moreover, ceding the debate about sensitive topics leaves a vacuum for uninformed, extremist, or unscientific perspectives to take hold (Reich, 2018).

Therefore, in this section of the article I present five lines of evidence that indicate that within the United States, average differences in intelligence between racial and ethnic groups are not 100%

environmental in origin. This evidence contradicts two viewpoints that are found in the literature. The first is that average differences in intelligence test scores across racial groups are purely environmentally caused (e.g., Helms, 1992; Ogbu, 2002). The second viewpoint is that there is no evidence supporting a genetic influence, so the position that average IQ differences are fully caused by environmental differences should be the default belief (e.g., Kaplan, 2015; Nisbett et al., 2012b; Sternberg, 2005). Scientists holding these viewpoints reach the same conclusion—that intergroup average differences in intelligence are environmentally caused—via two different perspectives. The first is more extreme and states that environmental differences can explain the entire mean IQ gap between racial groups; the latter is a moderate position that argues that genetic evidence has not been found (or cannot be) or that the genetic evidence simply does not exist. This article will review the evidence that both viewpoints are incorrect.

None of these lines of evidence is incontrovertible proof of a genetic influence of between-group differences in intelligence. However, the five together provide strong circumstantial evidence that points to the same conclusion: a nonzero genetic influence on racial and ethnic groups' mean intelligence differences. Although there are other types of evidence that some advocates of a genetic cause for mean group differences find convincing, I have limited myself to the evidence that I believe is most difficult to explain with a theory of purely environmental causes of average score gaps between demographic groups.

Before I describe these five lines of evidence, limits to this discussion need to be established. First, in this article I discuss only research on American populations. The reason for this is pragmatic: There has been more research on this topic in the United States than in any other nation. It may not be justifiable to generalize the research in this discussion to other countries or on a worldwide scale. Second, when members of different groups are discussed, it is assumed that each group consists solely of native English speakers born in the United States. This is because intelligence tests are designed for people who fit this demographic profile, and professionally designed intelligence tests are unbiased against these examinees (Jensen, 1980a; Reynolds, 2000). This eliminates from this article any discussion of language barriers, large cultural differences, or unfavorable environmental influences that originate in other nations (e.g., famine, civil unrest, a low-quality school system in immigrant examinees' nations of birth).

Note About Racial Groups and Classifications

It is important to understand the nature of racial groups to avoid misunderstandings or oversimplifications. A modern understanding from population genetics defines races as ancestral breeding populations that had their recent evolutionary origin in a particular geographic region of the world (Reich, 2018). These groups are defined in terms of the differences in the relative frequencies of a large number of alleles in the genome. These genetic differences arose because of relative isolation that the groups had from one another as natural selection or genetic drift occurred over the generations. As a result of this separation in recent evolutionary history, DNA variants can be used to identify the populations that a person's recent ancestors originated from. This information aligns almost perfectly with self-identified racial and ethnic group classifications. For example, in one typical study, atheoretical DNA-based classification matched self-identified race and ethnicity with 99.86% concordance (Tang et al., 2005). Results such as these should put to rest any claims that race is purely a social construct with no grounding in biology (see also Murray, 2020). Most of the research I discuss in this article uses self-identification to classify people into groups, although some of the genomic studies use DNA-based classification. Either method produces the same results for almost all individuals.

However, it is essential to recognize that these populations in recent evolutionary history are all the product of ancient admixtures of lineages that had previously split from the basal human population (Reich, 2018). Racial groups are not static, unchanging populations and never have been. Rather, current groups are identifiable at a particular point in human evolutionary history; analyses of DNA from ancient humans produces racelike groupings that do not align with modern racial groupings. These ancient genetic populations intermixed, branched off, or merged with other populations to produce modern racial groups. As a result, modern racial groupings make sense for contemporary populations, but there is nothing enduring or unchanging about any group, and there is no genotypic or phenotypic characteristic that all individuals within a racial group share and that is missing from all non–group members (Reich, 2018). However, none of this changes the fact that modern racial groups are identifiable via genomic analysis and that *average* genetic differences are easily observable when we compare sample of living humans who are descended from recent ancestral populations who lived in different regions of the world.

Evidence 1: Spearman's Hypothesis

Based on the approximate mean scores mentioned previously, the average intelligence difference between racial and ethnic groups in the United States ranges from about 5 to 20 IQ points, or .33 to 1.33 *SD* (Gottfredson, 1997a; Hunt, 2011), depending on which groups one chooses to compare. However, the magnitude of these differences varies from subtest to subtest, even within the same intelligence test battery (Jensen, 1980a). Psychologists have long recognized this fact (e.g., Derrick, 1920), although most did not investigate the issue. When discussing the varying mean score differences between African Americans and White Americans on subtests, Spearman (1927, pp. 379–380) off-handedly suggested that the better measures of *g* exhibited larger score gaps. This idea stayed untested until revived by Jensen (1980b), who named it Spearman's hypothesis.

Jensen (1985) created a test of Spearman's hypothesis, called the method of correlated vectors (MCV). In this procedure, the subtest's factor loading on a *g* factor—which is a measure of the degree to which the subtest measures *g*—and the effect size measuring the Black–White mean differences are correlated with one another. MCV results when comparing racial groups often show that these two vectors are positively correlated, indicating that Spearman was correct that better measures of *g* tend to have wider mean score differences between groups. These correlations vary but are consistently positive (e.g., $r = .84$ in Dahlke & Sackett, 2017; $r = .59$ in Jensen, 1985; $r = .46$ in Kane, 2007; $r = .57$ in te Nijenhuis & van den Hoek, 2016; $r = .723$ to $r = .817$ in Warne, 2016b). When expanded to other comparisons, the MCV seems to support Spearman's hypothesis when researchers compare White and Hispanic groups (Dahlke & Sackett, 2017; Hartmann, Kruuse, & Nyborg, 2007; Kane, 2007; te Nijenhuis, van den Hoek, & Dragt, 2019; Warne, 2016b) and White and Native American groups in the United States (te Nijenhuis, van den Hoek, & Armstrong, 2015). Results for Asian–White MCV analyses are inconsistent (Kane, 2007; Nagoshi, Johnson, DeFries, Wilson, & Vandenberg, 1984; Warne, 2016b).

The importance of Spearman's hypothesis for the study of causes of mean racial group differences in intelligence test scores is not obvious at first. However, if Spearman's hypothesis is true, then it would indicate that within- and between-group differences in intelligence arise from many of the same sources (genes and environment) because the sources of within- and between-group variation are positively correlated and share variance with one another. This finding would be consistent with the belief that group mean differences in intelligence are mostly (or perhaps entirely) the product of the accumulation of individual differences (Jensen, 1998; Rushton & Jensen, 2005). Moreover, it would indicate that any unique environmental influences that operate on one group would be trivial in their impact on between-group differences.

A strength of the Spearman's hypothesis literature is that it is a subset of a larger body of literature using the MCV. These studies consistently show that subtest *g* loadings correlate positively with other biological characteristics of the brain and nervous system (Jensen, 1998, Chapter 6). More tellingly, though, is the finding that the MCV produces positive correlations with phenomena that are purely genetic, such as inbreeding depression (Rushton, 1999), and zero correlations with phenomena that are purely environmental in origin, such as the Flynn effect (Rushton & Jensen, 2010), adoption (te Nijenhuis, Jongeneel-Grimen, & Armstrong, 2015), brain damage caused by trauma or prenatal exposure to toxins (Flynn, te Nijenhuis, & Metzen, 2014), and increased guessing on test items (Woodley, te Nijenhuis, Must, & Must, 2014). This finding provides evidence that positive correlations between *g* loadings and the magnitude of the gap in test scores between most groups in the United States probably has a partially genetic cause. MCV tests that produce zero correlations for purely environmentally caused group differences also contradict the claim that Spearman's hypothesis is a methodological artifact (e.g., Guttman, 1992).

However, it is important to recognize the limits of the MCV evidence regarding Spearman's hypothesis. One problem is that the MCV lacks specificity and that a positive correlation does not conclusively prove that mean differences between groups are due to differences in *g* or the same causes as differences in g. This shortcoming arises because the method does not detect violations of statistical assumptions that must be met in order to justify such conclusions (Dolan & Hamaker, 2001; Lubke, Dolan, & Kelderman, 2001; Wicherts, 2017). Additionally, MCV results are sensitive to subjective decisions about data preparation, such as the selection of subtests used to identify a *g* factor (Ashton & Lee, 2005), although collecting scores from a wide variety of subtest formats generally solves this problem (Jensen, 1992, 1998). Some statistical corrections are available to adjust for the restriction of range that can develop from a narrow selection of subtests (te Nijenhuis et al., 2019), although these rely on assumptions that may be difficult to prove. Thus, although the results of a positive correlation in an MCV analysis are consistent with Spearman's hypothesis and the belief that between- and within-group differences have the same causes, it is not conclusive evidence. Nonetheless, the logic of the MCV is attractive, and when data meet the necessary statistical assumptions and results are based on meta-analyses instead of single studies, Spearman's hypothesis provides evidence in support of the claim that mean group differences in intelligence are not fully environmental in origin.

Evidence 2: Tests of Measurement Invariance

In response to the inherent weaknesses of the MCV, some people in the intelligence research community began applying tests of measurement invariance to the study of group differences in intelligence. Measurement invariance was created by Meredith (1993) in order to strengthen the comparison of psychometric scores across demographic groups. The procedure uses multigroup confirmatory factor analysis (CFA) to test whether the necessary and sufficient conditions for comparing scores (the same factor structure, mean structure, and factor loadings for all groups) are present in the data derived from different groups. When these conditions are met, then mean differences in scores across demographic groups of examinees are due to actual differences in the construct being measured (Meredith, 1993). Since the invention of tests of measurement invariance, psychologists have applied the technique to intelligence test data many times. These results usually show that the factor structure and other psychometric properties of intelligence test data show a great deal of consistency across demographic groups (e.g., Beaujean, McGlaughlin, & Margulies, 2009; Li, Sano, & Merwin, 1996; Maller, 2000; Wicherts et al., 2004).

Dolan (2000) and his colleagues (Dolan & Hamaker, 2001; Lubke et al., 2001) realized that measurement invariance could overcome the problems inherent in the MCV. For example, measurement invariance procedures explicitly test the assumptions needed to justify the conclusion that between-group differences in *g* are due to the same causes as within-group differences in g, a characteristic that the MCV lacks (Dolan & Hamaker, 2001; Lubke et al., 2001). Moreover, tests of measurement invariance are an application of multigroup CFA (Meredith, 1993), unlike the MCV with its foundation in exploratory factor analysis. This means that the advantages of CFA, such as measures of model fit and the ability to test competing models, are intrinsic properties of tests of measurement invariance. This permits more complex and sophisticated investigations than are possible with the MCV and eliminates the shortcomings that can accompany practical applications of the MCV. For example, Frisby and Beaujean (2015) could test Spearman's hypothesis with a nonhierarchical factor structure—something that is not possible with the MCV.

More importantly, though, if measurement invariance holds across groups, then individual and mean group differences in a variable *must* be due to the same causes (Dalliard, 2014; Lubke, Dolan, Kelderman, & Mellenbergh, 2003). Therefore, tests of measurement invariance can eliminate the possibility that an influence on a phenotype operates on one group and not another. This is a stronger conclusion than is permitted by the MCV procedure. Research on intelligence test scores in the United States shows that measurement invariance holds across racial groups. Indeed, conducting tests of measurement invariance is typical practice in the development of professionally designed tests because it indicates the absence of test bias (Warne, Yoon, & Price, 2014), which test creators must eliminate to meet professional ethical standards for test development (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

To sum up, when measurement invariance holds, intelligence tests measure the same construct in the same way for different demographic groups. Measurement invariance also indicates that the same environmental and genetic causes of individual differences in intelligence within racial groups also determine between-group mean differences. However, it is important to note that this does not necessarily indicate that heritability values (whether within different groups or between groups) will be equal. Measurement invariance also does not conclusively prove that between-group heritability is greater than zero because mean environmental differences may be so large that between-group heritability could be zero. The discussion regarding Evidence 3 expounds this idea.

Evidence 3: Mathematical Relationship of Within- and Between-Group Differences

The degree of heritability of intelligence within various human groups and between groups is an empirical question. As stated previously, most heritability estimates fall between .20 and .80. Most of these estimates are derived from samples of White individuals, although in a recent meta-analysis, the average heritability of intelligence for White, African American, and Hispanic American samples was between .58 and .73 (Pesta et al., 2020, Table 3), indicating that heritability of IQ has a degree of stability across the largest racial groups within the United States.

Despite the heritability of intelligence within groups, many writers claim that within-group heritability provides no information about between-group heritability (e.g., Neisser et al., 1996), often using a thought experiment popularized by Lewontin (1970). In this analogy, one could take two random handfuls of seeds, planting one handful in fertile, moist soil. The other handful would be planted in barren, dry soil. After the seeds sprout and grow, any resulting phenotypic differences within either group must be completely genetic because the seeds within each group all experienced the same environment (i.e., all remaining differences are due solely to genetics). Conversely, all mean differences between groups must be solely environmental because the assignment of seeds to groups was random, and any genetic differences between groups have been balanced out.

Lewontin's analogy is flawed when applied to humans because humans are not randomly assigned to demographic groups (Warne et al., 2018). As a result, group mean differences in a phenotype probably are at least a partial consequence of aggregated individual differences in the phenotype. In turn, the individual differences are often partially caused by individual genetic differences.

DeFries (1972, p. 9) derived the mathematical relationship of between-group heritability and within-group heritability:

$$h_b^2 = (h_w^2)\frac{r_g(1-r_p)}{r_p(1-r_g)},$$

where

- h_b^2 is the heritability between groups (i.e., the proportion of between-group phenotype variance that is due to genetic differences between groups),
- h_w^2 is the heritability within groups,
- r_g is the genetic intraclass correlation within groups (which is a measure of genetic relatedness of group members), and
- r_p is the phenotypic intraclass correlation within groups (i.e., a measure of phenotypic similarity within a group).

The existence of the equation alone shows that the claim that "the genetic basis of the difference between two populations bears no logical or empirical relation to the heritability within populations" (Lewontin, 1970, p. 7) is incorrect. In the scenario of two randomly formed groups of organisms (or genetically identical organisms raised in separate environments), $r_g = 0$, which makes the entire right side of the equation equal to zero, indicating that $h_b^2 = 0$ for randomly formed groups (DeFries, 1972). This validates Lewontin's analogy; where Lewontin erred is in applying the special case of randomly formed groups to all human demographic groups, which by definition have r_g values greater than zero.[1]

Jensen (1998, pp. 447–458) applied DeFries's (1972) equation and showed that when $r_g > 0$, h_b^2 increases as h_w^2 increases, indicating that unless members of a racial group are completely unrelated to one another (i.e., $r_g = 0$, which by definition is impossible because racial group members belong to the same ancestral population), a value of h_b^2 greater than 0 is likely. However, an estimate of h_b^2 is not possible to derive because r_g is also unknown. Additionally, even with non–randomly formed groups, it is still possible for h_b^2 to be zero if between-group environmental differences are large enough to overpower the influence of between-group genetic differences (Warne, 2020b, Chapter 28).

---

[1] Incidentally, this discussion demonstrates the power of random assignment in true experiments. Random assignment of subjects to experimental and control groups balances out these groups genetically, which means that any later differences on the dependent variable(s) cannot be due to genetic differences between groups. Given the ubiquity of genetic influences on behavior (Plomin et al., 2016), random assignment is a powerful protection against genetic confoundin

The question then becomes whether the mean environmental differences between Black and White individuals in the United States are so large that they result in a between-group heritability (h_b^2) value of zero (Flynn, 1980). Expressed in standard deviation units, the required mean differences in environment associated with a between-group heritability (h_b^2) are expressed as Jensen's (1998, p. 454) equation:

$$\sigma_E = \sqrt{\frac{1-h_b^2}{1-h_w^2}},$$

where σ_E is the mean environmental difference between groups expressed in standard deviation units (i.e., Cohen's *d* effect size).

Whereas h_w^2 has empirically derived values, h_b^2 and σ_E remain unknown. Nevertheless, the equation can provide circumstantial evidence regarding the value of h_b^2. Armed with this equation, one can postulate various values for h_b^2 and h_w^2 and then examine whether the resulting σ_E values are plausible for a given mean between-group difference in a phenotype. Table 1 shows these calculations for a $d = 1.0$ difference in intelligence. Some of the σ_E values are inherently implausible. For example, if within-group heritability values of intelligence for African American and White American individuals are typically between 0.20 and 0.80, the values of σ_E in the top two rows (where $h_w^2 = 0.00$ and 0.10) and the bottom row (where $h_w^2 = 0.90$) are not realistic.

Table 1 illustrates an important feature of the relationship between between-group heritability and within-group heritability: When the two heritability values are equal, a 1-*SD* mean environmental difference produces a 1-*SD* difference in intelligence test scores. This is illustrated in the diagonal of 1.000 values, the cells that have equal h_b^2 and h_w^2 values. However, when $h_b^2 > h_w^2$, mean differences in environment must be less than the observed phenotype difference, as indicated by the σ_E values less than 1.000 above the diagonal. Conversely, when $h_b^2 < h_w^2$, then the mean environment differences must be greater than the mean phenotype difference, as indicated by the σ_E values greater than 1.000 below the diagonal. An implication from this table is that anyone who postulates that between-group heritability is less than within-group heritability must expect a larger mean environmental difference between groups than the mean difference observed in the actual phenotype. Additionally, the lower the postulated h_b^2 value, the larger environmental differences must be to produce the observed mean phenotype difference. Applied to intergroup intelligence differences, this implies that for between-group differences to be entirely genetic (i.e., for h_b^2 to be zero), the mean between-group environmental differences must be larger than the $d = 1.00$ mean difference observed between African American and White American individuals. Empirical data about the size of these differences can help judge the plausibility of different h_b^2 values, including whether h_b^2 could be equal to zero.

"Environment" is a catch-all term for any nongenetic effect on a phenotype (Hunt, 2011). Still, relevant environmental influences on intelligence differences (whether within or between groups) must be causal in nature. Therefore, any discussion of environmental influences can ignore irrelevant or noncausal environmental factors. Although this may seem like a banal insight, it does whittle down all environmental variables to a theoretically relevant set of variables that could affect a phenotype (Jensen, 1998). One of the most frequently suggested environmental causes of mean differences in intelligence among African Americans and White Americans is socioeconomic status (e.g., Nisbett et al., 2012a). Although socioeconomic status is by no means the only plausible environmental cause for mean differences in intelligence across American racial groups, it does produce a useful test case for the claim that between-group heritability of intelligence is zero. The U.S. Department of Education's National Center for Educational Statistics has made data publicly available from several representative longitudinal samples of children. The mean differences in socioeconomic status—a composite variable created by department personnel based on parental income, occupation, educational level, and (in one sample, the High School Longitudinal Study) the urbanicity of the child's high school—in the initial year of each study are shown in Table 2. The mean differences between African American and White American children's family socioeconomic status ranges from $d = 0.604$ to $d = 0.771$, with a weighted average of $d = 0.658$. All these values fall far short of the $d = 1.118$ to $d = 2.236$ environmental differences needed to produce a between-group heritability value of zero. Given within-group heritability of .20 to .80, these socioeconomic status differences account for only 7.3% to 47.6% of the environmental variance needed to produce an h_b^2 value of zero.[2]

Table 1. Projected Values of Mean Environmental Differences (σ_E), Given Between-Group Heritability (h_b^2) and Within-Group Heritability (h_w^2) Values for a Phenotype (IQ) Difference of d = 1.00

	h_b^2										
h_w^2	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
.00	1.000	0.949	0.894	0.837	0.775	0.707	0.632	0.548	0.447	0.316	0.000
.10	1.054	1.000	0.943	0.882	0.816	0.745	0.667	0.577	0.471	0.333	0.000
.20	1.118	1.061	1.000	0.935	0.866	0.791	0.707	0.612	0.500	0.354	0.000
.30	1.195	1.134	1.069	1.000	0.926	0.845	0.756	0.655	0.535	0.378	0.000
.40	1.291	1.225	1.155	1.080	1.000	0.913	0.816	0.707	0.577	0.408	0.000
.50	1.414	1.342	1.265	1.183	1.095	1.000	0.894	0.775	0.632	0.447	0.000
.60	1.581	1.500	1.414	1.323	1.225	1.118	1.000	0.866	0.707	0.500	0.000
.70	1.826	1.732	1.633	1.528	1.414	1.291	1.155	1.000	0.816	0.577	0.000
.80	2.236	2.121	2.000	1.871	1.732	1.581	1.414	1.225	1.000	0.707	0.000
.90	3.162	3.000	2.828	2.646	2.449	2.236	2.000	1.732	1.414	1.000	0.000

No theorist advocating for an h_b^2 value of zero would argue that the environmental impacts on intelligence are captured entirely by childhood socioeconomic status. The fact that the Cohen's d values in Table 2 are not large enough to result in an h_b^2 value of zero is not damning in any way to the claim that between-group differences in intelligence are fully environmental. It is likely that other environmental variables contribute to between-group mean differences in intelligence between African Americans and White Americans. However, the difficulty then becomes generating a list of variables that have a known causal impact on intelligence test scores and are nonredundant with socioeconomic status and any other environmental variables that cause between-group differences (Flynn, 1980). With the unimpressive results of experimental efforts to permanently raise IQ scores among people in industrialized countries (e.g., Lipsey, Farran, & Durkin, 2018; U.S. Department of Health & Human Services, 2012) and the large reduction in effect sizes in correlational studies after for genetic confounds were controlled for (e.g., Bouchard, Lykken, Tellegen, & McGue, 1996), finding environmental variables that fit these requirements and that can account for the remaining 52.4% to 92.7% of environmental variance needed to produce an h_b^2 value of zero seems implausible (Warne et al., 2018). If the correlation of socioeconomic status with intelligence is not due to a causal impact of socioeconomic status on intelligence, then the amount of variance that must be explained by other environmental variables for h_b^2 to be zero increases. In other words, it may not be enough for other environmental variables to explain 52.4% to 92.7% of variance in intelligence for h_b^2 to be zero.

---

[2] This estimate is derived by squaring the Cohen's d values in Table 2 and dividing them by the squared σ_E values in Table 1 for an h_b^2 value of zero and an h_w^2 value between .20 and .80. The resulting values of 7.3% and 47.6% are the minimum and maximum percentages of needed total environmental variance shared with socioeconomic status variance that the Cohen's d values in Table 2 could produce. As would be expected mathematically, higher percentages result from larger socioeconomic status differences in Table 2 and lower h_w^2 values in Table 1.

However, the average socioeconomic differences between African American and White American children can also help eliminate some of the heritability values in Table 1 from consideration. If one accepts the weighted average of $d = .658$ as the best estimate of observed mean socioeconomic status differences between African Americans and White Americans, then any combination of h_b^2 and h_w^2 values that produces a needed environmental difference less than $d = .658$ is inherently implausible because such environmental differences would produce mean IQ score differences larger than $d = 1.0$. in Table 1. This eliminates scenarios in which high values of between-group heritability are coupled with low values of within-group heritability, such as $h_b^2 = 0.70$ and $h_w^2 = 0.20$, which would need only an environmental difference of $d = 0.612$ to appear.

Table 2. Mean Differences in African American and White American Socioeconomic Status in Representative Samples of Schoolchildren

Dataset	Child grade at data collection	Mean difference (Cohen's *d*)
National Educational Longitudinal Study of 1988	*8*	0.625
Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999	Kindergarten	0.771
Education Longitudinal Study of 2002	10	0.606
High School Longitudinal Study of 2009	9	0.604
All samples (weighted average)		0.658

The information in Tables 1 and 2 and the mathematical constraints on the relationship that exists for between-group heritability and within-group heritability are powerful circumstantial evidence that the causes of mean differences in intelligence between African Americans and White Americans are not completely environmental. However, the information in this section is not exact enough to produce a likely value of between-group heritability.

There are two further difficulties for a purely environmental explanation for average IQ differences between races: measurement error and genetic influences on environment. Measurement error attenuates linear correlations and therefore requires that observed group differences in environment variables be larger than reported in Table 1 in order to reduce h_b^2 to zero. For example, if one believes that $h_w^2 = .50$, then the needed mean environmental difference between groups is $d = 1.414$ if environmental variables and intelligence are measured without error. However, if reliability for both variables is .90, then the needed observed difference becomes $d = \frac{1.414}{\sqrt{r_{xx'}r_{yy'}}} = \frac{1.414}{\sqrt{(0.90)(0.90)}} = \frac{1.414}{\sqrt{0.81}} = \frac{1.414}{0.90} = 1.571$. Lower reliability values will drive the needed observed effect size even higher. As a comparison, Herrnstein and Murray (1996, p. 598) reported a Cronbach's α value of .76 for their measure of socioeconomic status and at least .90 for their measure of intelligence. Using the equation above implies that with these reliability values, an observed mean environmental difference of $d = \frac{1.414}{\sqrt{(0.76)(0.90)}} = \frac{1.414}{\sqrt{0.684}} = \frac{1.414}{0.827} = 1.710$ between groups is needed to produce an h_b^2 value of zero when h_w^2 is .50 for a latent, error free measure of intelligence.

The second difficulty of genetic influences on environment arises from the fact that most people live in environments that are at least partially influenced by their genes, which means that even an "environmental" variable is often partially caused by genetics (Kendler & Baker, 2007; Plomin et al., 2016). Controlling for genetic confounds in environmental variables will correct the overestimated explanatory power of the environmental variables in correlational data and make it even more difficult for these variables (e.g., socioeconomic status) to explain between-group mean differences in intelligence.

Evidence 4: Data from GWASs

The first three forms of evidence are all what James Flynn (1980) classified as indirect forms of evidence regarding the possible influence of genes on intergroup average differences in intelligence. All three rely on statistical and methodological characteristics of data to show that purely environmental explanations of mean group differences are implausible. Though useful, these types of evidence do not provide evidence about the functioning of genes themselves. Instead, they focus on the observed patterns of data and use them to infer the probability of a genetic impact. In contrast, direct evidence examines the impact of genetics and is stronger than indirect evidence because it relies on fewer untested assumptions about the environment or genes and produces strong inferences about how group mean IQ differences would change if environments were equalized (Flynn, 1980).

A new form of direct evidence that between-group differences in intelligence are not fully environmental comes from GWASs. Currently, hundreds of genetic variants have been identified that correlate with intelligence test scores (Savage et al., 2018). The identification of DNA segments associated with intelligence would be interesting in its own right, but once these portions of DNA are known, it is possible to create an estimate of genetic influence (based on known relevant variants) on a particular person's phenotype. This estimate, called a polygenic score, is a variable created by assigning a weight to each variant of each relevant portion of DNA. Then, the weights for the variants in a particular individual's genome are summed to create an overall polygenic score for that individual (see Plomin & von Stumm, 2018, for a nontechnical explanation with a conceptual example). These polygenic scores correlate with intelligence test scores modestly ($r = .228$, $r^2 = .052$), but as more DNA variants associated with intelligence are identified these correlations will increase (Plomin & von Stumm, 2018; Savage et al., 2018). It is important to note that the phenotype variance explained by specific genes identified via GWAS are lower than the total within-group heritability identified through kinship studies (5.2% vs. 20–80%). This difference is called missing heritability and is probably due to current GWASs having sample sizes too small or lack the genetic diversity needed to detect rare genetic variants that can affect intelligence and to the limitations of current technology (Plomin & Deary, 2015; Plomin & von Stumm, 2018; Yang et al., 2015). Missing heritability occurs with many other phenotypes and is not unique to intelligence.

GWASs and polygenic scores provide the key to determining whether specific segments of DNA are associated with average intelligence score differences across groups because if the same DNA segments are identified in different racial groups as correlating with intelligence or if polygenic scores are correlated with intelligence test scores for different racial groups, then the most parsimonious conclusion is that some genes determine intelligence phenotypes for multiple racial groups. This is exactly the sort of direct evidence that Flynn (1980) stated would be more convincing than the indirect evidence based on statistical and methodological inferences I discussed earlier in this article.

Thus far, the majority of GWASs have been conducted on samples that consist mostly or entirely of individuals of European descent (Popejoy & Fullerton, 2016). This limitation creates a problem because genetic variants are not uniformly distributed across human populations, because of geographic or social separations during human evolutionary history (Reich, 2018) and genetic drift. As a result, genetic variants associated with intelligence in Europeans may not generalize to other racial groups (Domingue, Belsky, Conley, Harris, & Boardman, 2015). Additionally, because they are based on the statistical principles of regression, polygenic scores face an inevitable shrinkage in explained phenotype variance when applied to an independent sample.

Despite these challenges, there are indications that polygenic scores and GWAS results from Europeans can be partially generalized to other racial groups. In one study (Domingue et al., 2015), a polygenic score calculation method for adult educational attainment calculated with genetic Europeans' GWAS data were used to generate polygenic scores for African Americans, and these polygenic scores correlated with adult educational attainment $r = .11$ ($r^2 = .012$, $p = .004$; compare the genetic Europeans' correlation of $r = .18$, $r^2 = .032$, $p < .001$). Although adult educational attainment is not the same as intelligence, the two traits are positively correlated both at the phenotypic level (Damian, Su, Shanahan, Trautwein, & Roberts, 2015; Domingue et al., 2015) and at the genetic level via shared genes (Krapohl et al., 2014; Lee et al., 2018; Okbay et al., 2016). In fact, polygenic scores for educational attainment in adults predict phenotype intelligence test scores better than polygenic scores for intelligence do (Plomin & von Stumm, 2018). Thus, the fact that polygenic scores derived from individuals of European ancestry can generalize to African Americans—albeit with some shrinkage—is direct evidence that at least some of these genes influence group differences in mean intelligence test scores.

Further evidence comes from another study of applying GWAS data derived from European samples to other populations. Piffer (2015) used known frequencies of genetic variants that in published GWASs were associated in individuals descended from Europeans with better performance on intelligence tests or other measures of general cognitive functioning (e.g., executive functioning, educational attainment). He found that the frequency of these alleles in a nation's population was strongly correlated with estimated national mean IQ ($r = .91$). The positive correlation was maintained after the degree of genetic relatedness between populations was controlled for, a finding that Piffer (2019) later replicated. Moreover, a random set of alleles had a much weaker relationship with national-level mean IQ. Just as in the Domingue et al. (2015) study, this provides evidence that some of the alleles associated with higher intelligence in Europeans also have this relationship in other groups. It is likely that at least some of these genes are responsible for mean intelligence differences across racial groups.

It is important to recognize the limitations of this line of evidence. One is that, compared with other continents of origin, people with ancestors from Europe are not a very genetically diverse population (The 1000 Genomes Project Consortium, 2015), which may limit the ability of a GWAS to detect genetic variants associated with a phenotype if those alleles are comparatively rare in Europeans. Another problem is that polygenic scores and other GWAS results derived from Europeans are less generalizable to other populations as the degree of genetic relationships with Europeans decreases. As a result, applying results of GWASs to other populations sometimes produces nonsensical or incorrect results (Martin et al., 2017).

Nevertheless, the results of the Domingue et al. (2015) and Piffer (2015, 2019) studies are evidence that mean intelligence score differences across racial groups are not entirely environmental in origin. In the future, GWAS samples are likely to become more diverse because the health benefits of identifying alleles associated with phenotypes (e.g., for heart disease, diabetes) should be extended to all human populations (Popejoy & Fullerton, 2016; Reich, 2018). If this research progresses as it did for Europeans (which is likely), researchers conducting some GWASs will collect intelligence scores—or scores for highly correlated variables, such as educational attainment—from these diverse samples. This research will probably have two consequences. First, researchers will identify new genetic variants associated with intelligence, which will improve polygenic scores. Second, polygenic scores derived from more diverse samples will be more generalizable (Martin et al., 2017). These developments, already under way, will also provide stronger evidence of whether intergroup mean differences in intelligence are partially genetically caused.

Evidence 5: Admixture Studies

Another form of direct evidence that sheds light on whether between-group differences are fully environmental in origin is admixture studies. This method studies admixed populations, which are groups that are descended from ancestral populations that are native to different parts of the world (often different continents). In admixture studies, researchers capitalize on the individual variability in ancestry source within the admixed population by correlating the proportion of individuals' ancestry that originates from a particular continent with a phenotype score. If this correlation is not zero, then this is evidence that the trait has a degree of between-group genetic influence (Dalliard, 2014; Martin et al., 2017). Modern genomic-based admixture studies originated in medical research and have been successful at identifying health phenotypes that are associated with genetic heritage from different continents (e.g., Cheng et al., 2012; Flores et al., 2012; Kao et al., 2008).

In the United States, the largest admixed populations are Hispanic Americans and African Americans. Among Hispanic Americans, an average of 55–70% of ancestry originated in Europe (Bryc, Durand, Macpherson, Reich, & Mountain, 2015; Kirkegaard, Woodley of Menie, Williams, Fuerst, & Meisenberg, 2019), with the rest originating in the Americas and Africa. For African Americans, an average of 15–25% of ancestry originated in Europe, with the remaining ancestry originating almost entirely in Africa (Bryc et al., 2015; Jin et al., 2012). Within these groups of admixed individuals, there is a positive correlation ($\underline{r}$ = .23 to .30) between intelligence and the degree of European ancestry an individual possesses (Hu, Lasker, Kirkegaard, & Fuerst, 2019; Kirkegaard et al., 2019; Lasker, Pesta, Fuerst, & Kirkegaard, 2019; Warne, 2020a). This correlation cannot be explained by within-group discrimination based on skin color (Hu et al., 2019; Krieger, Sidney, & Coakley, 1998; Lasker et al., 2019), and controlling for socioeconomic status does not reduce the correlation to zero (Lasker et al., 2019). A purely environmental explanation of group differences in average intelligence has difficulty explaining the correlations between European ancestry because the correlations, by definition, have a genetic component.

Admixture studies of intelligence based on genetic assertations of ancestry are a recent phenomenon, and thus far only four have been published. However, with ancestral DNA tests becoming widespread and the growing push to incorporate genomic data into the social sciences (e.g., Murray, 2020; Plomin, 2018), more admixture studies will probably follow. In addition to replications of these early studies, future admixture studies on other groups will be an important theoretical test of the hypothesis of mean differences in intelligence across groups having a genetic influence because admixture studies lead to testable hypotheses about the possibility of a genetic influence on between-group average differences in intelligence. For example, Warne (2020a) extrapolated from admixture data in African Americans and Hispanic Americans to predict that individuals with European heritage admixed with ancestry from groups with higher average scores on intelligence tests, such as East Asians, should have a *negative* correlation between the proportion of their European ancestry and their intelligence test score. It is plausible that Warne's (2020a) hypothesis will be supported in such admixture tests; in one study of White Americans, individuals with ancestry from European Jews, the ethnic group with the highest average intelligence test score in the world, had higher polygenic scores for intelligence than sample members from other European groups (Dunkel, Woodley of Menie, Pallesen, & Kirkegaard, 2019). If admixture studies of European Jews and East Asians show a negative correlation between intelligence test scores and non-Jewish European ancestry, then it would indicate that people with high intelligence are more likely to have, on average, more genetic variants associated with higher intelligence, plus some environmental advantages (Warne, 2020a).

Still, correlation—even a genetic correlation—is not causation. One well-known objection to the statement that genetic correlations must have a genetic cause is from Jencks (1979), who stated that if a heritable trait is correlated with a proximal environmental cause of group differences, then it can result in a specious correlation between genes and an environmental outcome. Jencks proposed a thought experiment where children with red hair are systematically denied access to schooling; as a result, their academic performance would be correlated with the alleles for red hair, but that would not indicate that the alleles that encode for red hair have a causal impact on academic performance. However, no one has ever shown that such arbitrary environmental deprivations actually occur in 21st-century America. A hypothetical scenario where a particular gene–phenotype correlation was spurious does not mean that all—or any—such correlations in the real world are spurious. The parsimonious explanation for a correlation between genetic variations and a psychological phenotype is that genetic variance is a partial cause of phenotype variance. Scientists who postulate an interpretation that a particular correlation is an artifact of differential environments have the burden of proof of producing data that such environmental differences across genetic groups really do occur.

Less Conclusive Forms of Evidence

Readers should not take this discussion to imply that these are the only five forms of evidence regarding the potential genetic nature of between-group differences in intelligence. One early effort to address this issue was through transracial adoption studies where African American children were adopted into White families (e.g., Moore, 1986; Tizard, 1974). The most prominent and best designed of these studies was the Minnesota Transracial Adoption Study (MTAS), which started in the 1970s. In the first report (Scarr & Weinberg, 1976), African American and interracial (White and African American ancestry) children with a mean age of 6.2 years and adopted into White families at an average age of 18 months had modestly high IQ scores ($M = 106.3$, $SD = 13.9$). This average was above the national White IQ average (100) and only marginally lower than the mean IQ of White adoptees (111.5). These results seemed at the time to herald strong evidence against a genetic hypothesis for intergroup racial differences in intelligence. However, in the follow-up study, the adoptees' average IQs in adolescence had regressed toward their racial groups' mean. This was true for White adoptees ($M = 105.6$), interracial adoptees ($M = 98.5$), and African American adoptees ($M = 89.4$; Weinberg, Scarr, & Waldman, 1992). Thus, the evidence from the MTAS is ambiguous. Adoptees probably did receive an environmentally driven boost to their IQs, as indicated by the averages that were higher than each racial group's average IQ, which shows the importance of differences in home environments in average IQ gaps across racial groups in the United States. But the relative gap between African Americans' and White Americans' average IQs was still approximately 1 *SD* (16.2 IQ points in adolescence), and interracial children had average scores in between, all results in accordance with a model based on genetic influences. But these score differences could reflect the different preadoption experiences of the different groups. (For example, the African American adoptees were adopted later and had more preadoption placements than the other groups.) In a later discussion of the MTAS, the researchers acknowledged that the results did not permit strong inferences about the impact of either genetic or environmental influences (Waldman, Weinberg, & Scarr, 1994). I concur with this assessment.

With the best transracial adoption study producing ambiguous evidence (and having methodological shortcomings that, in retrospect, weakened any possible conclusions), it is clear that the current transracial adoption literature is not productive to analyze when discussing the likelihood of genetic influences on average differences in intelligence within the United States. The MTAS and other transracial adoption studies were noble efforts, but because they lack methodological characteristics necessary to demonstrate unconfounded environmental or genetic effects, their evidence is ambiguous. Their most problematic shortcoming is an absence of intelligence test scores for the adopted children's biological parents. A transracial adoption study with intelligence test scores for biological and adopted parents, children from every large American minority group, and data on the environment created by both sets of parents (including the prenatal environment) would provide direct evidence that is stronger than the data that I have presented in this article. Unfortunately, no such study currently exists, and there have been no reported American transracial adoption studies on intelligence since the MTAS. There are some studies of transracial adopted children published in other countries, but these studies usually involve international adoptions. None report intelligence test scores for biological parents of adopted children, and few have detailed data about the preadoption environment (see van IJzendoorn, Juffer, & Klein Poelhuis, 2005, for a review).

Another piece of evidence that is sometimes used to argue that average intelligence differences are not fully environmentally caused is the data on brain size (e.g., Rushton & Jensen, 2005). Although it is true that brain volume correlates positively with IQ and that East Asians have larger brains than Europeans, who have larger brains than Africans (all these differences being in averages, with large amounts of overlap between individuals in various groups), I do not find this evidence convincing for three reasons. First, biological differences are not necessarily genetic in origin (Gottfredson, 2009), and it has not been conclusively shown that interracial differences in mean brain size have a genetic cause and that this genetic cause also has a causal impact on the correlation between brain size and IQ. Second, the advantage for individuals with larger brains is not apparent when comparing male and female subjects; male brains are larger than female brains, even when body size is controlled for, but both sexes have the same average intelligence (Jensen, 1998; Jensen & Johnson, 1994). This finding is problematic for the argument that big brains are smarter brains. Finally, evidence is mounting that brain organization and functioning may be more important biological determinants of intelligence than sheer brain size (Haier, 2017).

Other sources of evidence (see Rushton & Jensen, 2005, for a summary) are ambiguous in their support for either the fully environmental or partially genetic theories of the causes of average group differences in intelligence. This is why I have limited myself to the five sources of evidence I discuss for the bulk of this article.

Discussion

In this article I have presented five lines of indirect and direct evidence that mean group differences in intelligence in the United States are not 100% environmentally caused: findings in support of Spearman's hypothesis, consistent results from tests of measurement invariance across American racial groups, the mathematical relationship for between-group and within-group sources of heritability, the preliminary results of applying polygenic scores and the results of GWASs to non-European groups, and the results of admixture studies. No one line of evidence is incontrovertible proof that between-group differences in intelligence in the United States are partially genetic in origin (although the direct evidence is strongest). However, all five types of evidence point in the same direction: Racial group differences in intelligence are at least partially genetically caused. Meanwhile, there is no fully environmental theory that parsimoniously explains all the facts I have presented. Those wanting to adhere to a fully environmental explanation have so far constructed a series of ad hoc explanations, which lack the parsimony of a partially genetic explanation for the phenomena I describe (Gottfredson, 2005; Rushton & Jensen, 2005).

Moreover, it is not easy to see how current theories of fully environmental differences will accommodate the new direct evidence that is likely to emerge from the GWASs or future admixture studies, given the results that these methods have produced so far. On the other hand, a partially genetic explanation not only can accommodate such findings but can generate new, testable predictions about the relationship between genes, environment, and intelligence phenotypes across populations. For example, my belief that polygenic scores from diverse GWAS samples will explain more phenotype variance in intelligence and be generalizable to non-European samples is strongly dependent on the assumption that between-group differences in intelligence are at least partially genetic. If this hypothesis is disproved, then it would be a major setback to the claim of a genetic influence on intergroup intelligence differences. The same is true of my (Warne, 2020a) hypothesis regarding intelligence being positively correlated with East Asian heritage in European–Asian admixed groups.

Despite the diverse psychometric and genetic data I have presented, one nagging question remains unanswered: What percentage of between-group variance in intelligence scores is due to genetic variance between racial and ethnic groups in the United States? None of the evidence I have presented in this article provides a conclusive answer to this question. The percentage is almost certainly greater than zero, but beyond that, these sources of information do not point to an exact estimate for between-group heritability. There is no scientific reason why this question remains unanswered in the 21st century. The methods and technology exist now to determine the between-group heritability for racial groups in the United States. I believe there are at least two reasons for this gap in knowledge: a remarkable lack of curiosity among scientists and the lack of funding to answer such a question.

Studies of racial differences in economic, health, educational, behavioral, and other outcomes are ubiquitous in the social sciences. That discrepancies in many outcomes exist is beyond doubt, and there is no shortage of hypothesized environmental causes. Yet researchers who study these differences rarely postulate genetic causes. Even when genetic causes of discrepancies are raised, some scientists concentrate on denying such possibilities instead of confronting the data and testing the hypotheses surrounding potential genetic causes for interracial group differences in intelligence (e.g., Kwate, 2001; Lewontin, 1970; Sternberg, 2005). For a group that is supposed to be curious about the world, the number of scientists forming and testing hypotheses about the possibility of a genetic influence on racial differences in intelligence is surprisingly small. One reason may be the lack of ideological diversity in the social sciences, with overwhelmingly left-of-center political beliefs among researchers (Inbar & Lammers, 2012; Kaufman, 2021; Redding, 2001) that may cause some social scientists to not investigate the genetics of group differences because they find the topic distasteful (e.g., Horgan, 2013; Sternberg, 2005) or believe they already know the answer (see Flynn, 2012, p. 36, for a discussion of this latter reason). If ideological conformity is discouraging research on average race differences in intelligence, then it is an impediment to scholarly progress because it prevents the creation of knowledge that emerges from the Hegelian clash of contradicting viewpoints—the thesis and antithesis—to arrive at a more correct understanding of facts: the synthesis.

Another impediment to determining the degree to which genetic influences affect intergroup differences in intelligence is the lack of funding. Government agencies and private foundations shy away from funding controversial topics, and the genetics of racial differences in intelligence is certainly one of those topics. In one recent case, recounted by mathematics researcher Theodore P. Hill (2018), he and Sergei Tabachnikov submitted an article on variability differences in intelligence between sexes as a possible explanation for the preponderance of men in the highest and lowest levels of academic accomplishment. Much to Hill's surprise, the National Science Foundation (NSF) asked the authors to remove the acknowledgement of Hill's NSF funding from the article, ostensibly because the topic was unrelated to Tabachnikov's funding. However, emails that Hill obtained from a Freedom of Information Act request showed that the real reason was pressure on the NSF from the chair of the Climate and Diversity Committee and the associate head for diversity and equity at Tabachnikov's university (T. P. Hill, 2018). If a government agency bowed to such political pressure regarding a simple acknowledgment of funding on an article about a topic that is uncontroversial among intelligence researchers,[3] the pressure to not award grants to researchers to study race differences in intelligence must be overwhelming. For some granting organizations, such an award may even be unthinkable. The excuse that there are not sufficient funds for such research is unlikely. As James Flynn—no friend to the hypothesis of genetic causes of interracial differences in intelligence—stated,

> If universities have their way, the necessary research will never be done. They fund the most mundane research projects, but never seem to have funds to test for genetic differences between races. I tell US academics I can only assume that they believe that racial IQ differences have a genetic component, and fear what they might find. They never admit that the politics of race affects their research priorities. It is always just far more important to establish whether squirrels enjoy *The Magic Flute*. (Flynn, 2012, p. 36)

The aversion to investigating, funding, and promoting research on the genetic influences on average intelligence across racial groups has made the topic languish and has limited the amount of data available. As a result, the controversy has festered far longer than necessary, and it is difficult to arrive at exact conclusions based on strong empirical data. A full exploration of the causes of this dynamic (which are probably rooted in the sociology of science) is beyond the scope of this article. However, some scholars have postulated about causes of this phenomenon. Coleman (1991) suggested that a concern throughout the social sciences toward historically marginalized groups has led to an ethic of "conspicuous benevolence," where findings that reflect unfavorably on these groups go quietly unreported and studies that might reveal such findings are not conducted. Coleman believed that this ethic would lead to self-suppression of scholarly research in order for scientists to retain the approval of their colleagues, which is important for obtaining tenure, a promotion, favorable ratings for grant applications, or professional eminence.

---

[3] See Feingold (1992) for a landmark literature review on the topic. The finding of greater male variability is now widely accepted, with many different tests usually showing a 5–15% greater standard deviation on male cognitive test scores (e.g., Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Hur, te Nijenhuis, & Jeong, 2017; Lakin, 2013; Strand, Deary, & Smith, 2006).

A recent survey supports Coleman's (1991) view, finding that 43% of American academics and 30% of British academics in the humanities and social sciences stated that they would support the firing of a colleague who published work showing that hiring more minorities and women is correlated with lower organizational performance (Kaufman, 2021, p. 24). This scenario engendered more opposition than any of the other four controversial topics in the survey. A different survey of undergraduate students found that 85% of respondents would oppose a speaker on campus who claimed that "Some racial groups are less intelligent than others" (College Pulse, Foundation for Individual Rights in Education, & RealClear Education, n.d., p. 55). These surveys shows that Coleman (1991) may have been too mild; academics may not be avoiding race differences in IQ solely to retain approval from their colleagues, as important as that may be. They may be trying to keep their jobs or avoid being the center of a firestorm. These strong negative reactions from students and faculty seem to occur despite the data that indicate that a partially genetic cause of group differences in IQ is plausible.

Other reasons for the aversion toward researching group differences in IQ have been suggested. Davis (1978) believed that a "moralistic fallacy" was at work, where people's moral desires for what they want to be true stand in place for actual scientific truth—a cognitive heuristic that leads people to make decisions about scientific discovery accordingly. Humphreys (1988, 1991) suggested several widespread folk beliefs among social scientists that would discourage research on the possibility of a genetic influence on mean intelligence differences, including a bias among social scientists in favor of environmental causes of human behavior, a narrow focus on socioeconomic status as a powerful cause of individual and group differences in life outcomes, and a default belief that there could not be any genetic influences on group differences. Regardless of the cause, the result is that the topic of genetic influences on intergroup intelligence averages has been understudied, which allows the controversy to linger on unnecessarily for decades and prevents social scientists from crafting policy proposals based on sound data.

Conclusion

The five types of circumstantial evidence presented here are not, individually, conclusive regarding the potential of genetic influences on mean differences among racial groups in the United States. Each has its own strengths and weaknesses. For example, the GWAS evidence consists of just three studies, but these studies provide consistent direct evidence of the influence of specific genes on intelligence in different racial groups. The measurement invariance evidence cannot provide an estimate of the degree to which between-group variance is genetic, but it does eliminate the possibility of a unique influence that uniformly lowers one group's intelligence test scores while leaving another group untouched (see Warne, 2020b, Chapters 29–30, for a further discussion of the unlikelihood of such unique influences). Combined, all five sources of data make a cohesive argument that group differences in intelligence are not fully environmentally caused.

I call upon psychologists to have an open mind and to investigate the evidence for themselves, starting with the sources I have cited in this article. I also encourage social scientists to make research contributions that can address this question. If readers believe in fully environmental causes for mean racial group differences in intelligence, I invite them to design their own studies to test their hypothesis and to create a coherent theory of environmental causes of intelligence that can parsimoniously explain the findings I have summarized and make new, testable predictions. Controversy thrives in ignorance; it is time to move beyond the basic question of whether genes have an impact on intergroup mean IQs so that scientists can investigate productive questions of the magnitude of this influence and how to accommodate differences in a diverse society.

References

1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68–74. https://doi.org/10.1038/nature15393

Adams, H. H. H., Hibar, D. P., Chouraki, V., Stein, J. L., Nyquist, P. A., Rentería, M. E., . . . Thompson, P. M. (2016). Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nature Neuroscience, 19*, 1569–1583. https://doi.org/10.1038/nn.4398

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Arden, R., Luciano, M., Deary, I. J., Reynolds, C. A., Pedersen, N. L., Plassman, B. L., . . . Visscher, P. M. (2016). The association between intelligence and lifespan is mostly genetic. *International Journal of Epidemiology, 45,* 178–185. https://doi.org/10.1093/ije/dyv112

Asbury, K., Moran, N., & Plomin, R. (2016). Nonshared environmental influences on academic achievement at age 16: A qualitative hypothesis-generating monozygotic-twin differences study. *AERA Open, 2*(4). https://doi.org/10.1177/2332858416673596

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence, 33,* 431–444. https://doi.org/10.1016/j.intell.2004.12.004

Beaujean, A. A., McGlaughlin, S. M., & Margulies, A. S. (2009). Factorial validity of the Reynolds Intellectual Assessment Scales for referred students. *Psychology in the Schools, 46,* 932–950. https://doi.org/10.1002/pits.20435

Binet, A., & Simon, T. (1905/1916). New methods for the diagnosis of the intellectual level of subnormals (E. S. Kite, trans.). In A. Binet & T. Simon, *The development of intelligence in children (the Binet–Simon Scale)* (pp. 37–90). Williams & Wilkins.

Bouchard, T. Jr. (2014). Genes, evolution and intelligence. *Behavior Genetics, 44,* 549–577. https://doi.org/10.1007/s10519-014-9646-x

Bouchard, T. J. Jr., Lykken, D. T., Tellegen, A., & McGue, M. (1996). Genes, drives, environment, and experience: EPD theory revised. In C. P. Benbow (Ed.), *Intellectual talent: Psychometric and social issues* (pp. 5–43). Johns Hopkins University Press.

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics, 96,* 37–53. https://doi.org/10.1016/j.ajhg.2014.11.010

Burks, B. S. (1973). The relative influence of nature and nurture upon mental development: A comparative study of foster parent–foster child resemblance and true parent–true child resemblance. In H. J. Eysenck (Ed.), *The measurement of intelligence* (pp. 325–357). Springer. (Original work published in 1928)

Cheng, C.-Y., Reich, D., Haiman, C. A., Tandon, A., Patterson, N., Selvin, E., . . . Kao, W. H. L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: A genetic admixture analysis in three U.S. Population cohorts. *PLOS ONE, 7*(3), e32840. https://doi.org/10.1371/journal.pone.0032840

Cofnas, N. (2016). Science is not always "self-correcting": Fact–value conflation and the study of intelligence. *Foundations of Science, 21*(3), 477–492. https://doi.org/10.1007/s10699-015-9421-3

Coleman, J. S. (1991). The Sidney Hook Memorial Award address: On the self-suppression of academic freedom. *Academic Questions, 4*(1), 17–22. https://doi.org/10.1007/BF02682943

College Pulse, Foundation for Individual Rights in Education, & RealClear Education. (n.d.). *2020 College Free Speech Rankings: What's the climate for free speech on America's college campuses?* https://www.thefire.org/research/publications/student-surveys/2020-college-free-speech-rankings/

Curran, F. C., & Kellogg, A. T. (2016). Understanding science achievement gaps by race/ethnicity and gender in kindergarten and first grade. *Educational Researcher, 45,* 273–282. https://doi.org/10.3102/0013189x16656611

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology, 102,* 1403–1420. https://doi.org/10.1037/apl0000234

Dalliard, M. (2014). The elusive x-factor: A critique of J. M. Kaplan's model of race and IQ. *Open Differential Psychology*. https://doi.org/10.26775/ODP.2014.08.25

Damian, R. I., Su, R., Shanahan, M., Trautwein, U., & Roberts, B. W. (2015). Can personality traits and intelligence compensate for background disadvantage? Predicting status attainment in adulthood. *Journal of Personality and Social Psychology, 109,* 473–489. https://doi.org/10.1037/pspp0000024

Davis, B. B. (1978). The moralistic fallacy. *Nature, 272*(5652), 390. https://doi.org/10.1038/272390a0

Deary, I. J. (2012). 125 years of intelligence in *The American Journal of Psychology. American Journal of Psychology, 125,* 145–154. https://doi.org/10.5406/amerjpsyc.125.2.0145

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews: Neuroscience, 11,* 201–211. https://doi.org/10.1038/nrn2793

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence, 31,* 533–542. https://doi.org/10.1016/s0160-2896(03)00053-9

DeFries, J. C. (1972). Quantitative aspects of genetics and environment in the determination of behavior. In L. Ehrman, G. S. Omenn, & E. Caspari (Eds.), *Genetics, environment, and behavior* (pp. 5–16). Academic Press.

Derrick, S. M. (1920). A comparative study of the intelligence of seventy-five whites and fifty-five colored college students by the Stanford revision of the Binet–Simon scale. *Journal of Applied Psychology, 4,* 316–329. https://doi.org/10.1037/h0071332

Detterman, D. K. (2006). Editorial note on controversial papers. *Intelligence, 34,* iv. https://doi.org/10.1016/j.intell.2005.10.001

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17,* 913–920. https://doi.org/10.1111/j.1467-9280.2006.01802.x

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research, 35,* 21–50. https://doi.org/10.1207/S15327906MBR3501_2

Dolan, C. V., & Hamaker, E. L. (2001). Investigating Black–White differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC and a critique of the method of correlated vectors. In F. H. Columbus (Ed.), *Advances in psychology research* (Vol. 6, pp. 31–59). Nova Science Publishers.

Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., & Boardman, J. D. (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA Open, 1*(3), 1–13. https://doi.org/10.1177/2332858415599972

Dunkel, C. S., Woodley of Menie, M. A., Pallesen, J., & Kirkegaard, E. O. W. (2019). Polygenic scores mediate the Jewish phenotypic advantage in educational attainment and cognitive ability compared with Catholics and Lutherans. *Evolutionary Behavioral Sciences, 13,* 366–375. https://doi.org/10.1037/ebs0000158

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62,* 61–84. https://doi.org/10.3102/00346543062001061

Flashman, L. A., Andreasen, N. C., Flaum, M., & Swayze, V. W., II. (1997). Intelligence and regional brain volumes in normal controls. *Intelligence, 25*(3), 149-160. https://doi.org/10.1016/S0160-2896(97)90039-8

Flores, C., Ma, S.-F., Pino-Yanes, M., Wade, M. S., Pérez-Méndez, L., Kittles, R. A., . . . Garcia, J. G. N. (2012). African Ancestry is associated with asthma risk in African Americans. *PLOS ONE, 7*(1), Article e26807. https://doi.org/10.1371/journal.pone.0026807

Flynn, J. R. (1980). *Race, IQ, and Jensen*. Routledge & Kegan Paul.

Flynn, J. R. (2012). *Are we getting smarter? Rising IQ in the twenty-first century*. Cambridge University Press.

Flynn, J. R., te Nijenhuis, J., & Metzen, D. (2014). The g beyond Spearman's g: Flynn's paradoxes resolved using four exploratory meta-analyses. *Intelligence, 44,* 1–10. https://doi.org/10.1016/j.intell.2014.01.009

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence, 51,* 79–97. https://doi.org/10.1016/j.intell.2015.04.007

Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Hossiep, R., Voelkle, M. C., . . . Jung, R. E. (2018). Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. *Nature Communications, 9,* 1905. https://doi.org/10.1038/s41467-018-04268-8

Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence, 64,* 18–29. https://doi.org/10.1016/j.intell.2017.06.004

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24,* 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Gottfredson, L. S. (1997b). Why g matters: The complexity of everyday life. *Intelligence, 24,* 79–132. https://doi.org/10.1016/S0160-2896(97)90014-3

Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law, 11,* 311–319. https://doi.org/10.1037/1076-8971.11.2.311

Gottfredson, L. S. (2009). Logical fallacies used to dismiss the evidence on intelligence testing. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 11–65). American Psychological Association.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13,* 1–4. https://doi.org/10.1111/j.0963-7214.2004.01301001.x

Gould, S. J. (1981). *The mismeasure of man*. W. W. Norton.

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research, 27,* 175–204. https://doi.org/10.1207/s15327906mbr2702_2

Haier, R. J. (2017). *The neuroscience of intelligence*. Cambridge University Press.

Hartmann, P., Kruuse, N. H. S., & Nyborg, H. (2007). Testing the cross-racial generality of Spearman's hypothesis in two samples. *Intelligence, 35,* 47–57. https://doi.org/10.1016/j.intell.2006.04.004

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47,* 1083–1101. https://doi.org/10.1037/0003-066X.47.9.1083

Herrnstein, R. J., & Murray, C. (1996). *The bell curve: Intelligence and class structure in American life* (2nd ed.). Free Press.

Hill, T. P. (2018, September 7). Academic activists send a published paper down the memory hole. *Quillette. https://quillette.com/2018/09/07/academic-activists-send-a-published-paper-down-the-memory-hole/*

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., . . . Deary, I. J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry, 24,* 169–181.

Horgan, J., (2013, May 16). Should research on race and IQ be banned? [Web log post] *Scientific American Cross-Check Blog.* https://blogs.scientificamerican.com/cross-check/should-research-on-race-and-iq-be-banned/

Hu, M., Lasker, J., Kirkegaard, E. O. W., & Fuerst, J. G. R. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych, 1,* 240–261. https://doi.org/10.3390/psych1010017

Humphreys, L. G. (1988). Trends in levels of academic achievement of blacks and other minorities. *Intelligence, 12,* 231–260. https://doi.org/10.1016/0160-2896(88)90025-6

Humphreys, L. G. (1991). Limited vision in the social sciences. *American Journal of Psychology, 104,* 333–353. https://doi.org/10.2307/1423243

Hunt, E. (2011). *Human intelligence*. Cambridge University Press.

Hur, Y.-M., te Nijenhuis, J., & Jeong, H.-U. (2017). Testing Lynn's theory of sex differences in intelligence in a large sample of Nigerian school-aged children and adolescents (n > 11,000) using Raven's Standard Progressive Matrices Plus. *Mankind Quarterly, 57*, 428–437.

Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science, 7,* 496–503. https://doi.org/10.1177/1745691612448792

Jencks, C. (1979). *Who gets ahead? The determinants of economic success in America*. Basic Books.

Jensen, A. R. (1980a). *Bias in mental testing*. The Free Press.

Jensen, A. R. (1980b). Précis of bias in mental testing. *Behavioral and Brain Sciences, 3,* 325–333. *https://doi.org/10.1017/S0140525X00005161*

Jensen, A. R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8,* 193–219. https://doi.org/10.1017/S0140525X00020392

Jensen, A. R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research, 27,* 225–233. https://doi.org/10.1207/s15327906mbr2702_5

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

Jensen, A. R., & Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence, 18*, 309–333. https://doi.org/10.1016/0160-2896(94)90032-9

Jin, W., Xu, S., Wang, H., Yu, Y., Shen, Y., Wu, B., & Jin, L. (2012). Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Research, 22,* 519–527. https://doi.org/10.1101/gr.124784.111

Jung, R. E., & Haier, R. J. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences, 30,* 135–154. https://doi.org/10.1017/S0140525X07001185

Kane, H. (2007). Group differences in nonverbal intelligence: Support for the influence of Spearman's g. *Mankind Quarterly, 48,* 65–82.

Kao, W. H. L., Klag, M. J., Meoni, L. A., Reich, D., Berthier-Schaad, Y., Li, M., . . . Parekh, R. S. (2008). MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nature Genetics, 40*, 1185–1192. https://doi.org/10.1038/ng.232

Kaplan, J. M. (2015). Race, IQ, and the search for statistical signals associated with so-called "X"-factors: Environments, racism, and the "hereditarian hypothesis." *Biology & Philosophy, 30,* 1–17.

Kaufman, E. (2021, March 1). *Academic freedom in crisis: Punishment, political discrimination, and self-censorship* (Report no. 2). Center for the Study of Partisanship and Ideology. https://cspicenter.org/wp-content/uploads/2021/03/AcademicFreedom.pdf

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: A systematic review. *Psychological Medicine, 37,* 615–626. https://doi.org/10.1017/S0033291706009524

Kirkegaard, E. O. W., Woodley of Menie, M. A., Williams, R. L., Fuerst, J., & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych, 1,* 1–25. https://doi.org/10.3390/Psychology1010001

Kourany, J. A. (2016). Should some knowledge be forbidden? The case of cognitive differences research. *Philosophy of Science, 83,* 779–790. https://doi.org/https://doi.org/10.1086/687863

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., . . . Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111,* 15273–15278. https://doi.org/10.1073/pnas.1408777111

Krieger, N., Sidney, S., & Coakley, E. (1998). Racial discrimination and skin color in the CARDIA study: Implications for public health research. *American Journal of Public Health, 88,* 1308–1313. https://doi.org/10.2105/ajph.88.9.1308

Kwate, N. O. A. (2001). Intelligence or misorientation? Eurocentrism in the WISC-III. *Journal of Black Psychology, 27,* 221–238. https://doi.org/10.1177/0095798401027002005

Lakin, J. M. (2013). Sex differences in reasoning abilities: Surprising evidence that male–female ratios in the tails of the quantitative reasoning distribution have increased. *Intelligence, 41,* 263–274.

https://doi.org/10.1016/j.intell.2013.04.004

Lasker, J., Pesta, B. J., Fuerst, J. G. R., & Kirkegaard, E. O. W. (2019). Global ancestry and cognitive ability. *Psych, 1,* 431–459. https://doi.org/10.3390/psych1010034

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics, 50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lewontin, R. C. (1970). Race and intelligence. *Bulletin of the Atomic Scientists, 26*(3), 2–8. https://doi.org/10.1080/00963402.1970.11457774

Li, X., Sano, H., & Merwin, J. C. (1996). Perception and reasoning abilities among American, Japanese, and Chinese adolescents. *Journal of Adolescent Research, 11,* 173–193. https://doi.org/10.1177/0743554896112002

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45,* 155–176. https://doi.org/10.1016/j.ecresq.2018.03.005

Little, M. (2017). Racial and socioeconomic gaps in executive function skills in early elementary school. *Educational Researcher, 46,* 103–109. https://doi.org/10.3102/0013189x17698700

Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating group differences on cognitive tests using Spearman's hypothesis: An evaluation of Jensen's method. *Multivariate Behavioral Research, 36,* 299–324. https://doi.org/10.1207/s15327906299-324

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence, 31,* 543–566. https://doi.org/10.1016/S0160-2896(03)00051-5

Maller, S. J. (2000). Item invariance in four subtests of the Universal Nonverbal Intelligence Test (UNIT) across groups of deaf and hearing children. *Journal of Psychoeducational Assessment, 18,* 240–254. https://doi.org/10.1177/073428290001800304

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., . . . Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *American Journal of Human Genetics, 100,* 635–649. https://doi.org/10.1016/j.ajhg.2017.03.004

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543. https://doi.org/10.1007/BF02294825

Mollon, J., Knowles, E. E. M., Mathias, S. R., Gur, R., Peralta, J. M., Weiner, D. J., . . . Glahn, D. C. (2021). Genetic influence on cognitive development between childhood and adulthood. *Molecular Psychiatry, 26,* 656–665.

Moore, E. G. (1986). Family socialization and the IQ test performance of traditionally and transracially adopted Black children. *Developmental Psychology, 22,* 317–326. https://doi.org/10.1037/0012-1649.22.3.317

Murray, C. (2007). The magnitude and components of change in the black–white IQ difference from 1920 to 1991: A birth cohort analysis of the Woodcock–Johnson standardizations. *Intelligence, 35,* 305–318. https://doi.org/10.1016/j.intell.2007.02.001

Murray, C. (2020). *Human diversity: The biology of gender, race, and class*. Twelve.

Murray, C. (2021). *Facing reality: Two truths about race in America*. Encounter Books.

Nagoshi, C. T., Johnson, R. C., DeFries, J. C., Wilson, J. R., & Vandenberg, S. G. (1984). Group differences and first principal-component loadings in the Hawaii family study of cognition: A test of the generality of "Spearman's hypothesis." *Personality and Individual Differences, 5,* 751–753.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012a). Group differences in IQ are best understood as environmental in origin. *American Psychologist, 67*(6), 503-504. https://doi.org/10.1037/a0029772

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012b). Intelligence: New findings and theoretical developments. *American Psychologist, 67,* 130–159. https://doi.org/10.1037/a0026699

Ogbu, J. U. (2002). Cultural amplifiers of intelligence: IQ and minority status in cross-cultural perspective. In J. M. Fish (Ed.), *Race and intelligence: Separating science from myth* (pp. 241–278). Erlbaum.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539–542. https://doi.org/10.1038/nature17671

Pesta, B. J., Kirkegaard, E. O. W., te Nijenhuis, J., Lasker, J., & Fuerst, J. G. R. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence, 78*, Article 101408. https://doi.org/10.1016/j.intell.2019.101408

Penke, L., Maniega, S. M., Bastin, M. E., Valdes Hernandez, M. C., Murray, C., Royle, N. A., . . . Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry, 17,* 1026–1030. https://doi.org/10.1038/mp.2012.66

Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence, 53,* 43–50. https://doi.org/10.1016/j.intell.2015.08.008

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from Gwas hits: A replication of previous findings using recent data. *Psych, 1,* 55–75. https://doi.org/10.3390/psych1010005

Plomin, R. (2018). *Blueprint: How DNA makes us who we are*. MIT Press.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry, 20,* 98–108. https://doi.org/10.1038/mp.2014.105

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11,* 3–23. https://doi.org/10.1177/1745691615617439

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148–159. https://doi.org/10.1038/nrg.2017.104

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature, 538*(7624), 161–164. https://doi.org/10.1038/538161a

Redding, R. E. (2001). Sociopolitical diversity in psychology: The case for pluralism. *American Psychologist, 56,* 205–215

Reich, D. (2018). *Who we are and how we got here: Ancient DNA and the new science of the human past.* Pantheon Books.

Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law, 6,* 144–150. *https://doi.org/10.1037/1076-8971.6.1.144*

Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression—unlike Black–White differences: A reply to Flynn. *Personality and Individual Differences, 26,* 381–389.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11,* 235–294.

Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect a narrowing of the Black–White IQ gap. *Intelligence, 38,* 213–219. https://doi.org/10.1016/j.intell.2009.12.002

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., . . . Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50*, 912–919.

Scarr, S., & Weinberg, R. A. (1976). IQ test performance of Black children adopted by White families. *American Psychologist, 31*, 726–739.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201–292. https://doi.org/10.2307/1412107

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. Macmillan.

Sternberg, R. J. (2005). There are no public-policy implications: A reply to Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11,* 295–301.

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in Cognitive Abilities Test scores: A UK national picture. *British Journal of Educational Psychology, 76,* 463–480. https://doi.org/10.1348/000709905x50906

Tang, H., Quertermous, T., Rodriguez, B., Kardia, S. L. R., Zhu, X., Brown, A., . . . Risch, N. J. (2005). Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *American Journal of Human Genetics, 76*(2), 268–275. https://doi.org/10.1086/427888

te Nijenhuis, J., Jongeneel-Grimen, B., & Armstrong, E. L. (2015). Are adoption gains on the g factor? A meta-analysis. *Personality and Individual Differences, 73,* 56–60. https://doi.org/10.1016/j.paid.2014.09.022

te Nijenhuis, J., & van den Hoek, M. (2016). Spearman's hypothesis tested on black adults: A meta-analysis. *Journal of Intelligence, 4*, 6. https://doi.org/10.3390/jintelligence4020006

te Nijenhuis, J., van den Hoek, M., & Armstrong, E. L. (2015). Spearman's hypothesis and Amerindians: A meta-analysis. *Intelligence, 50*, 87–92. https://doi.org/10.1016/j.intell.2015.02.006

te Nijenhuis, J., van den Hoek, M., & Dragt, J. (2019). A meta-analysis of Spearman's hypothesis tested on Latin-American Hispanics, including a new way to correct for imperfectly measuring the construct of g. *Psych, 1*, 101–122. https://doi.org/10.3390/psych1010008

Tizard, B. (1974). IQ and race. *Nature, 247*(5439), 316. https://doi.org/10.1038/247316a0

Turkheimer, E. (2007). Race and IQ. Cato Unbound. https://www.cato-unbound.org/2007/11/21/eric-turkheimer/race-iq

U.S. Department of Health and Human Services. (2012). Third grade follow-up to *Head Start impact study. Final report* (OPRE report 2012-45).

van IJzendoorn, M. H., Juffer, F., & Klein Poelhuis, C. W. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin, 131,* 301–316. https://doi.org/10.1037/0033-2909.131.2.301

Waldman, I. D., Weinberg, R. A., & Scarr, S. (1994). Racial-group differences in IQ in the Minnesota Transracial Adoption Study: A reply to Levin and Lynn. *Intelligence, 19,* 29–44. https://doi.org/10.1016/0160-2896(94)90051-5

Warne, R. T. (2016a). Five reasons to put the g back into giftedness: An argument for applying the Cattell–Horn–Carroll theory of intelligence to gifted education research and practice. *Gifted Child Quarterly, 60,* 3–15. https://doi.org/10.1177/0016986215605360

Warne, R. T. (2016b). Testing Spearman's hypothesis with advanced placement examination data. *Intelligence, 57,* 87–95. https://doi.org/10.1016/j.intell.2016.05.002

Warne, R. T. (2020a). Continental genetic ancestry source correlates with global cognitive ability score. *Mankind Quarterly, 60,* 400–422.

Warne, R. T. (2020b). *In the know: Debunking 35 myths about human intelligence*. Cambridge University Press. https://doi.org/10.1017/9781108593298

Warne, R. T., Astle, M. C., & Hill, J. C. (2018). What do undergraduates learn about human intelligence? An analysis of introductory psychology textbooks. *Archives of Scientific Psychology, 6,* 32–50. https://doi.org/10.1037/arc0000038

Warne, R. T., Yoon, M., & Price, C. J. (2014). Exploring the various interpretations of "test bias." *Cultural Diversity & Ethnic Minority Psychology, 20,* 570–582. https://doi.org/10.1037/a0036503

Weinberg, R. A., Scarr, S., & Waldman, I. D. (1992). The Minnesota Transracial Adoption Study: A follow-up of IQ test performance at adolescence. *Intelligence, 16,* 117–135. https://doi.org/10.1016/0160-2896(92)90028-p

Wicherts, J. M. (2017). Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence, 60,* 26–38. https://doi.org/10.1016/j.intell.2016.11.002

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32,* 509–537. https://doi.org/10.1016/j.intell.2004.07.002

Woodley, M. A., te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the Brand effect. *Intelligence, 43,* 27–34. https://doi.org/10.1016/j.intell.2013.12.004

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., . . . Visscher, P. M. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics, 47*(10), 1114–1120. https://doi.org/10.1038/ng.3390

PART 2:

ADMIXTURE REGRESSION ANALYSIS OF RACE-RELATED TEST SCORE GAPS

CHAPTER 4

CONTINENTAL GENETIC ANCESTRY SOURCE CORRELATES WITH GLOBAL COGNITIVE ABILITY SCORE

RUSSELL T. WARNE

Abstract

The sources of between-group variation in cognitive ability has been a source of controversy for decades. In this pre-registered study, I tested the hypothesis that such differences are entirely environmental by examining the correlation between the proportion of individuals' genetic heritage that originates on different continents with their total cognitive test score. In a sample of 193 Hispanic Americans and African Americans, it was found that the proportion of European genetic heritage was positively correlated ($r = .277$ to .300, $p<.001$) with the total cognitive test score. This makes any assertion of a purely environmental cause of mean group differences in cognitive ability less probable. One possible explanation is that the ancestors of Europeans experienced stronger selection pressures for high cognitive ability than the ancestral groups for Africans and Native Americans. However, this study may or may not generalize to other groups or a worldwide sample. Molecular genetics data suggest that most genetic causes are not specific for European (or any other continental) heritage. Such genetic variations are polymorphic in most human populations. Any person — from any demographic group — with a high proportion of such alleles is more likely than others to have high general cognitive ability.

Keywords: cognitive ability, genetic admixture, genetic ancestry

1. Introduction

General cognitive ability has been shown to be highly heritable in European-origin populations in industrialized countries, with experts typically stating that heritability for general cognitive ability is .40 to .70 (e.g., Bouchard, 2014; Plomin et al., 2016). Some researchers even find heritability of cognitive ability in some samples exceeding .80 (e.g., Jacobs et al., 2001). One limitation of heritability studies is that they can only report the estimated heritability for the studied population in its current environment. Heritability studies in White populations in industrialized countries tell scientists little about the heritability of cognitive ability in other racial/ethnic groups living under different conditions, or about genetic and environmental contributions to differences between nations (Bouchard, 2014). As a result, some have claimed that heritability studies of European populations and their descendants elsewhere in the world provide little or no insight into the heritability within other groups (e.g., Scarr-Salapatek, 1971).

Likewise, several researchers — the most prominent of whom was Lewontin (1970) — have argued that high heritability of intragroup individual differences says nothing about genetic or non-genetic origins of between-group differences (e.g., Herrnstein, 1971; Herrnstein & Murray, 1996; Neisser et al., 1996; Thoday, 1969). These scholars claim that even high heritability within White populations can theoretically coincide with completely environmental causes of between-group differences in mean cognitive ability levels. When examining potential genetic causes of racial and ethnic group differences, some scientists argue that there is little or no evidence for genetic causes (e.g., Neisser et al., 1996; Nisbett et al., 2012). On the other hand, a majority of experts on human cognitive abilities believe that there is a nonzero genetic contribution to differences in cognitive ability between African Americans and European Americans (Rindermann, Becker & Coyle, 2020; Snyderman & Rothman, 1987) and internationally (Rindermann, Becker & Coyle, 2016).

There are a variety of indirect ways that one can test whether between-group differences are purely environmental, including tests of measurement invariance (e.g., Lubke et al., 2003), investigations of Spearman's hypothesis (e.g., Dahlke & Sackett, 2017; Jensen, 1985; te Nijenshuis & van den Hoek, 2016; Warne, 2016), and biometric decomposition models to estimate the needed mean differences in environment to cause known mean score gaps (e.g., Jensen, 1998; Warne, in press). However, some scientists have proposed a direct way of testing whether between-group differences are purely environmental in origin is to investigate a genetically admixed population and determine whether there is a correlation between the proportion of genetic heritage from a high-ability population and individuals' intellectual ability scores (Dalliard, 2014; Lee, 2010; Rushton & Jensen, 2005).

Dalliard's (2014) proposal is not new; such studies occurred through much of the 20th century. Witty and Jenkins (1936) and Shuey (1966, Chapter 9) reported literature reviews of studies examining the degree of White admixture in African Americans. The authors found that there were inconsistent correlations between African Americans' amount of European ancestry (provided by self-report or estimated by the degree of skin pigmentation) and general mental ability. Later researchers used blood group markers to estimate degree of European ancestry among African Americans. Loehlin, Vandenberg and Osborne (1973, pp. 264-267) found in two samples of African Americans (*N*s = 20 and 22 twin pairs) correlations of $r = -.38$ and $r = .01$ between continental ancestry estimated from blood marker frequency and cognitive ability, indicating that European ancestry has either no relationship with general cognitive ability, or a negative relationship. However, neither correlation was statistically significant. In a similar study, Scarr et al. (1977, p. 82) found a correlation of $r = -.05$ between estimated European admixture based on blood group markers and cognitive test performance. Likewise, in a study of Chilean Hispanics (Rothhammer & Llop, 1976, p. 460) estimated Native American admixture derived from blood group markers was correlated $r = .02$ with cognitive ability scores. The results of these studies indicated a lack of correlation between estimated European admixture and cognitive ability.

The early methods of determining genetic admixture are now obsolete. Blood group-based methods used a small number of traits (16 or fewer) to determine the degree of admixture in a person's ancestry. However, accuracy of estimates increases as the number of traits examined increases (Novembre & Peter, 2016; Smouse, Spielman & Park, 1982), and modern DNA-based methods use thousands of portions of a person's genome to generate highly accurate estimates of individuals' degree of genetic heritage from different populations (see Mersha & Abebe, 2015, for an accessible explanation). At the individual level, these results correlate highly with self-report racial classification (Bryc et al., 2015; Lima-Costa et al., 2015; Tang et al., 2005). At the group level, the results coincide with the historical movements of populations (Abdellaoui et al., 2019; Bryc et al., 2015; Han et al., 2017), evolutionary history (Tang et al., 2005), and with the documented degrees of coexistence of population groups within a geographic area (Adhikari et al., 2017; Bryc et al., 2015).

From an evolutionary perspective, there is good reason to believe that general cognitive ability for an individual would be influenced by the degree of admixture in their ancestry. First, after subgroups of early humans left Africa, they and their descendants were exposed not only to population bottlenecks that changed allele frequencies in unpredictable ways, but also to different ecological environments — and therefore different selection pressures. Some theorists have posited that cognitive ability may be one of the traits selected for in some groups over time (e.g., Cochran, Hardy & Harpending, 2006; Cochran & Harpending, 2009; León & Burga-León, 2015; Rindermann, 2018; Rushton, 2000; Unz, 2013; Winegard, Winegard & Boutwell, 2017; Woodley of Menie et al., 2017). Indeed, the evidence is strong that human evolution occurred at a faster rate in the past 50,000 years than at any time before anatomically modern humans left Africa (Cochran & Harpending, 2009). If the theory of accelerated evolution of human traits — including cognitive ability for some populations — were true, then one would expect that the tens of thousands of years of exposure to different environments and evolutionary pressures would produce among different human populations differing frequencies of alleles associated with high cognitive ability. This plausibly has resulted in differing mean cognitive ability levels across human subpopulations.

However, in the last 500 years or so, large-scale cross-continental migration and human intermixing have increased. This has resulted in more human populations that are recently admixed — i.e., descended from two or more evolutionarily distinct parent populations. Yet, within any recently admixed population there is variability in the proportions of genes that individuals inherited from the original parent populations. If racial group differences in cognitive ability are at least partially genetically caused, then there should be a positive correlation between (a) individuals' proportion of genes from the parent population with a mean higher cognitive ability and (b) admixed individuals' cognitive ability. Conversely, a zero correlation (or negative correlation) would be evidence against the hypothesis that genetic differences contribute to between-group differences in cognitive ability (Nisbett, 2009, p. 97).

Some researchers have engaged in preliminary investigations of this hypothesis at the group level. Piffer (2013) found that the Met allele of the functional catechol-O-methyltransferase (COMT) Val158Met polymorphism (which is associated with higher executive functioning and working memory) was more common worldwide in agricultural populations — who usually display higher cognitive abilities — than in hunter-gatherer groups, whereas the pattern of prevalence of the Val allele was reversed. Researchers have also found that alleles associated with cognitive ability in the FNBP1L gene (rs236330) (Piffer, 2013), MCPH1, and ASPM or MCPH5 (Woodley et al., 2014) are more prevalent in populations with higher mean cognitive ability.

Using a different methodology—but with data more commonly available for human groups—it is apparent that in western hemisphere nations, the proportion of citizens with ancestries originating in Europe is positively correlated with the nation's mean cognitive ability level. Conversely, the proportion of citizens with ancestries originating in the Americas or Africa is negatively correlated with national average cognitive ability (Fuerst & Kirkegaard, 2016). Within the same article, these researchers replicated the results at the state level for Mexico, Brazil, and Colombia; in later articles they replicated the results in the United States (Kirkegaard & Fuerst, 2016) and Argentina (Kirkegaard & Fuerst, 2017).

While results at the group level are informative, there is the possibility that they may fall prey to the ecological fallacy. In other words, it is possible that group-level results may not apply at the individual level. The purpose of the current study is to update and expand upon prior research on the correlation between the degree of European admixture in American minority populations and cognitive ability by using modern methods to ascertain the proportion of individuals' ancestry that originates on different continents. In this study, the sample consisted of African Americans and Hispanic Americans. These groups were chosen because they both possess a high degree of genetic admixture. African Americans possess, on average, 10-30% European DNA (Adhikari et al., 2017; Bryc et al., 2015; Fuerst & Kirkegaard, 2016; Jin et al., 2012; Mersha & Abebe, 2015; Tishkoff et al., 2009), whereas American Hispanics have an average of 15-73% non-European ancestry (Adhikari et al., 2017; Bryc et al., 2015; Kirkegaard & Fuerst, 2016). This high level of genetic admixture within each population should provide sufficient variation in genetic ancestry for any correlations with phenotypes to be observable. Prior research has shown that the ancestry sources in admixed populations can correlate with health-related phenotypes (Mersha & Abebe, 2015). Therefore, it seems plausible that ancestry variables can correlate with a psychological phenotype.

2. Methods

2.1. Sample

Individuals for the study were recruited online via Qualtrics Panels, an online commercial data collection service which is similar to Amazon MTurk. To be eligible, individuals had to self-report that they were Hispanic or African American and that they had access to the results of at least one commercial ancestry DNA test. Individuals who did not meet these requirements or did not provide informed consent were immediately removed from the study.

The original intended sample size was 178 African Americans and 178 Hispanic Americans, for a total of 356 individuals (the maximum sample size possible, given the project budget). Within subgroups of this size, a one-tailed test of statistical significance would have 80% power if the correlation were $r = .186$. For the two groups combined, the same power can be achieved with a correlation as small as $r = .132$. In the initial stages of data collection it was found that the vast majority of respondents were not eligible for the study because they had not taken an ancestry DNA test. Among those who claimed they had taken such a test, a high percentage of individuals (50-70%) provided unusable responses. After six rounds of data collection, the staff at Qualtrics and I came to the conclusion that it would be extremely difficult to increase the sample size further. At the end of data collection, the final sample size was 193. This still allowed for statistical power of 80% if the correlation between variables in the entire sample was $r = .183$. The intention to analyze the Hispanic and African American subsamples separately was abandoned at this point.

2.2. Materials and Stimuli

With the innovation of commercial ancestry DNA testing, millions of people in the United States have access to high quality estimates of the origins of their genetic ancestry (Han et al., 2017). At the beginning of the online data collection process, individuals were asked to access the results of their commercial ancestry test. (Individuals who had taken more than one of these tests were asked to access their most recent test.) They were then asked to type in the results, with the instruction to ignore "Low Confidence Regions" (for ancestry.com customers) or populations that contributed less than 4.5% of the person's DNA (for 23andMe and other companies' customers) because these results are less accurate and sometimes unstable. For a person's response to be considered valid, they had to provide percentages that summed to a value between 60% and 100% and consisted of only the names of regions of the world (e.g., "Africa North," "Iberian Peninsula," "Benin") or ancestral populations (e.g. "Native American"). Responses that were considered invalid included "Hispanic," "Mexican," "African American," and "American."

After providing their genetic ancestry test results, subjects completed a short 27-item cognitive test created by the International Cognitive Ability Resource team (2014). This test consisted of four subtests: 3D Rotation, Letter-Number Series (a pattern recognition task), Matrix Reasoning, and Verbal Reasoning. Three of the tests (all except Letter-Number Series) were shortened by only administering half of the items.[1] This was done to reduce the chance of examinee fatigue. A description of the validity evidence that the test measures general cognitive ability can be found in Condon and Revelle (2014).

2.3. Data Preparation and Statistical Analysis

Reported ancestry percentages were summed to produce a percentage of each individual's ancestry that originated in (a) Europe, (b) Africa, (c) the Americas, (d) Asia, and (e) other parts of the world. The last two groups were combined into one category because only one individual reported ancestry from other parts of the world. All percentages were converted to proportions. Cognitive ability test items were scored dichotomously (1 point for a correct answer and 0 points for an incorrect answer), and item scores were summed to produce a score for each subtest and a total score for the entire test.

Descriptive statistics (means and standard deviations) were calculated for each continent's proportion of genetic ancestry (Africa, Europe, the Americas, and Asia/other parts of the world) and the cognitive test score. Cronbach's α was calculated to estimate each subtest score's and the total test score's internal consistency reliability. The Pearson's r value was used to estimate the correlation between the proportion of each continental source of genetic heritage and the total cognitive test score. These correlations were tested for statistical significance ($\alpha = .05$). A one-tailed test was chosen (and pre-registered) because Europeans typically outscore Africans and Native Americans on these tests and because at the group level mean European ancestry proportion was the strongest predictor of mean cognitive ability (Fuerst & Kirkegaard, 2016; Kirkegaard & Fuerst, 2016, 2017). Therefore, it seemed likely that if there were a relationship between European ancestry and cognitive test scores, the correlation would be positive.

2.4. Study Pre-Registration

The methods for this study were pre-registered through the Open Science Framework (OSF) at https://osf.io/bc5fv/. The site also provides a copy of the ethical approval for the study, the data file, and the SPSS analysis syntax and output. Only the cognitive test items are not reproduced on the OSF site because these are confidential.

3. Results

3.1. Descriptive Statistics

Table 1 reports the demographic data of the sample members. The sample was 73.1% female and—unsurprisingly, given the study participation requirements—55.4% African American and 47.2% Hispanic. On average, the respondents reported a .319 proportion of European ancestry, .359 proportion of African ancestry, 0.191 proportion of Native American ancestry, and .020 proportion of Asian/other ancestry.

---

[1] The pre-registration protocol states that all tests would be shortened by administering half of the items. This was a mistake in the pre-registration; the Letter-Number series subtest is only 9 items long and can be answered rapidly. It was never intended to be shortened.

Table 1: Sample demographics ($N = 193$) and descriptive statistics

Variable	# (%)	Mean	SD	Cronbach's α
Males	52 (26.9%)	—	—	—
Females	141 (73.1%)	—	—	—
African Americans[a]	107 (55.4%)	—	—	—
Asian Americans[a]	1 (0.5%)	—	—	—
Hispanics/Latinos[a]	91 (47.2%)	—	—	—
Native Americans/Alaska Natives[a]	9 (4.7%)	—	—	—
Pacific Islanders[a]	1 (0.5%)	—	—	—
Whites/Europeans[a]	35 (18.1%)	—	—	—
Other[a]	1 ("Irish," 0.5%)	—	—	—
European ancestry proportion	—	0.319	0.287	—
African ancestry proportion	—	0.359	0.344	—
Native American ancestry proportion	—	0.191	0.252	—
Other ancestry proportion	—	0.020	0.074	—
3D Rotation test (6 items)	—	0.750	1.067	.533
Letter-Number Series test (9 items)	—	3.920	2.689	.806
Matrix Reasoning test (6 items)	—	2.070	1.560	.568
Verbal Reasoning test (6 items)	—	3.350	1.283	.463
Total test score (27 items)	—	10.090	4.893	.823

[a] Racial/ethnic group numbers exceed 193 because respondents were allowed to choose multiple categories.

Table 1 also reports the mean and standard deviation for scores on the cognitive skills subtests and the total test score. Cronbach's α values were low for three out of the four subtests (with Letter-Number series being the exception, likely because it is the longest subtest). The overall test score Cronbach's α value was .823.

3.2. Correlations

Table 2 shows the correlations among the proportion of individuals' ancestry that originates in Europe, Africa, the Americas, and the rest of the world and the total test score. European ancestry proportion was moderately correlated with total test scores (r = .297, p<.001). Although the correlations between the total test score and the proportions of African (r = -.132, p = .033) and Native American (r = -.142, p = .024) ancestry were statistically significant, readers must realize that these p-values are based on pre-registered one-tailed correlations. Neither would have been statistically significant under a two-tailed test. Though these correlations are not high, it is apparent that there is enough variation within the sample to show a slight negative correlation between African or Native American ancestry proportions and cognitive ability. But a larger sample size is needed for these correlations to stabilize enough to permit strong conclusions about the true strength of the relationship between these ancestry proportions and cognitive ability.

Table 2: Correlation matrix of ancestry proportions and total test score

	EUR	AFR	NTV	OTH
European ancestry proportion (EUR)	1.000			
African ancestry proportion (AFR)	-.665**	1.000		
Native American ancestry proportion (NTV)	-.121*	-.565**	1.000	
Other ancestry proportion (OTH)	-.111	-.034	-.069	1.000
Total cognitive ability test score (COG)	.297**	-.132*	-.142*	-.039

Although not pre-registered, two analyses were conducted to determine whether the correlation between European ancestry proportion and total cognitive test score would vary under different data preparation procedures. First, reported ancestry proportions almost always did not sum to 1.0 for respondents — a common occurrence because some individuals' genome contains portions that cannot be confidently assigned to an ancestral population (Bryc et al., 2015) and because respondents were told to ignore small percentages in their ancestry report. Therefore, I adjusted these values so that any unknown proportion of a respondent's ancestry would be distributed proportionately among European, African, Native American, and Asian/other ancestry groups. Thus, if a person stated that their ancestry was 60% African and 30% European, the remaining 10% would be apportioned so that the adjusted percentages would be 66.7% African ancestry and 33.3% European. This would ensure that all participants' ancestry proportions would sum to 1.0. Second, a principal axis factor analysis was conducted on the subtest scores to produce a *g* factor (which explained 51.2% of variance), and a *g* score was calculated for each participant. These changes did not produce substantial changes in the correlation between European ancestry proportion and total cognitive test score ($r = .277$ to .300, all $ps < .001$),[2] probably because the two measures of general cognitive ability were highly correlated ($r = .994$), as were the reported and adjusted ancestry proportions ($r = .979$ to .993).

Another non-pre-registered analysis was suggested by a *Mankind Quarterly* peer reviewer. At his or her behest, I created a multiple regression model that had the cognitive ability test score (converted to *z*-scores) as a dependent variable and the European ancestry proportion and Hispanic dummy coded variables (both unstandardized) as independent variables. The results are shown in Table 3. The results indicate that controlling for race/ethnicity self-identification does little to change the relationship between European ancestry and total test score. The unstandardized regression (b) value for the European ancestry variable means that the predicted cognitive ability test score difference for two people—one with zero European ancestry and another with 100% European ancestry—is 1.078 standard deviations. With the standardized *g* score as the dependent variable (table not shown, but available at https://osf.io/bc5fv/), the b value was .912 for the European ancestry independent variable; Hispanic self-identification had no statistical or practical significance in the model.

Table 3: Regression model with European ancestry proportion and Hispanic dummy code independent variables and cognitive ability test score dependent variable

Independent variable	b	Standard error	β	P
	Model 1			
Constant	-.330	.103	—	.002
European ancestry proportion	1.034	.240	.297	<.001
	Model 2			
Constant	-.324	.107	—	.003
European ancestry proportion	1.078	.295	.310	<.001
Hispanic dummy code	-.043	.170	-.022	.800

---

[2] These analyses are not shown in the article but are available on the OSF page.

4. Discussion

4.1. Summary

This study was an attempt to find a correlation between European ancestry and general cognitive test scores in American racial minority group members. Research in the 20th century was often unable to demonstrate a correlation between these two variables (e.g., Loehlin, Vandenberg & Osborne, 1973; Rothhammer & Llop, 1976; Scarr et al., 1977; Witty & Jenkins, 1936). This study shows that such a correlation is detectable with contemporary methods of estimating an individual's proportion of ancestry from various areas of the world. The results indicate that admixture studies with larger sample sizes would likely be fruitful in finding relationships between continental genetic sources and cognitive ability scores.

The results of this study support the results from group-level analyses of the relationship between genetic ancestry and cognitive ability. For example, when correlating mean admixture levels for nations and states in the Americas, a group's proportion of European ancestry is consistently a stronger predictor of mean cognitive ability than the proportion of African or Native American ancestry (Fuerst & Kirkegaard, 2016; Kirkegaard & Fuerst, 2016, 2017). Likewise, these results support Piffer's (2013) argument that many alleles for high (or low) cognitive ability were probably selected for in conjunction with one another and that these alleles are likely to be more frequent in some human populations than others.

One implication of this study is that the hypothesis of purely environmental causes of between-group mean differences in general intellectual ability is less probable. Given these admixture results, the most parsimonious explanation is that mean differences among groups today are partially genetically caused. It implies that the ancestors of Europeans experienced greater selection pressures for high cognitive ability in their recent evolutionary history than the ancestors of Africans or the peoples of the Americas.

This study shows a direct connection between intergroup genetic variation and cognitive test scores at the individual level, indicating that at least some intergroup variance is likely genetically caused (Dalliard, 2014). Purely environmental theories of mean group differences in cognitive ability must plausibly explain how and why these environmental variables mimic genetic influences. Additionally, if environmental influences can closely imitate the effects of genetic differences, then scientists who favor purely environmental explanations of group differences must explain what type of evidence would disprove their hypothesis (Gottfredson, 2005).

4.2. Independent Confirmation

I conducted this study in early 2018. Unbeknownst to me when I started the study, another independent research group was investigating the same research question with larger datasets. This group published two studies while my article was under consideration at other journals. In the first study (Kirkegaard et al., 2019), which consisted of data from a sample of $N = 1{,}369$ and exact ancestral heritage data (instead of relying on self-report), this other research group found that their entire sample (41.4% White, 23.6% Hispanic, 10.1% African American, and 24.9% other/multiracial), European ancestry proportion was correlated with cognitive ability at $r = .23$ (two-tailed $p<.001$; Kirkegaard et al., 2019, Table 6). Within subsamples of Hispanic and African Americans, the correlation between cognitive ability and European ancestry proportion was $r = .18$ to $.24$ (two-tailed ps $= .0001$ to $.035$; Kirkegaard et al., 2019, Tables S9-S11). Though these values are slightly weaker than the correlations in my data, they are all within the 95% confidence intervals for my study's sample size ($N = 193$) and correlations ($r = .277$ to $.300$). The second study (Lasker et al., 2019) showed a correlation of $r = .251$ between European ancestry proportion and cognitive ability scores in a sample of 2,407 monoracial African Americans and biracial European/African Americans. There is no statistically significant difference between the correlations in all three of these studies, and each serves as a replication of the others. Additionally, the other studies controlled for self-identified racial group membership, socioeconomic status, and non-native English speaker status (Kirkegaard et al., 2019; Lasker et al., 2019). Thus, these potentially confounding variables likely do not explain the relationship found in these studies between European ancestry proportion and cognitive ability.

4.3. Limitations

There are shortcomings of this study that readers should be aware of. First, individuals' genetic ancestry results were not independently verified. It is possible that some results are inaccurate or incomplete.

Second, when results were converted into an analyzable format, proportions from ancestral populations on the same continent were summed so that the results were simplified to four variables. However, some continents (e.g., Asia and Africa) are so genetically diverse that it may not make sense to combine these proportions into one continental ancestry proportion (Warne, 2016). As an example, Sub-Saharan Africans seem to be a distinct genetic clade from north Africans (Tishkoff et al., 2009), the latter of whom are much more closely related to Europeans. However, in this study I combined north African ancestry (which many Hispanic subjects reported) and Sub-Saharan ancestry because both originate in Africa and I wanted to reduce subjectivity in the ancestry calculations. At the behest of a *Mankind Quarterly* peer reviewer, I reapportioned the North African ancestry to the European group and found that the correlations between European ancestry and cognitive ability scores were still very similar ($r = .218$ for the raw sum scores and $r = .213$ for the g scores, both one-tailed $p\text{s} = .0015$).

It is also important to note that this study consisted of a convenience sample. However, the same is true for all prior studies of the relationship between genetic ancestry and cognitive ability. My sample size was also larger than most studies that used blood groups to estimate genetic ancestry. Rothhammer and Llop (1976) studied 63 Chilean college students, while Loehlin et al. (1973) collected data from two samples of Black children totaling 84 individuals in 42 twin pairs. Only Scarr et al. (1977), Lasker et al. (2019), and Kirkegaard et al. (2019) exceeded this study's sample size. Still, a larger sample size would be preferable, and I have made my data available to other researchers who wish to combine it with other admixture data for future meta-analyses. Another issue with the sampling of this study is that there is no reason to believe that customers of a commercial ancestry DNA service are representative of larger racial or ethnic groups that these individuals belong to (Kirkegaard & Fuerst, 2016).

Finally, this study's measure of cognitive ability consisted of four short subtests, each 6-9 items long. While the overall score's reliability was adequate for research purposes ($\alpha = .823$), reliability was too low for most of the subtests to determine whether ancestry was correlated with specific cognitive task scores. Such information could provide insight into the mental traits that are most closely related to genetic ancestry, which would have important implications in educational achievement gaps across groups (e.g., Warne, 2016) or in differences in job success and/or placement (Roth et al., 2001).

It is interesting, though, that these shortcomings made it harder to detect a correlation between ancestry and cognitive ability in this study because the procedures should introduce measurement error into the variables. Yet, a statistically significant relationship between European ancestry and cognitive ability still emerged. With more accurate and reliable measures of the variables, the correlations shown in Table 2 may increase, though this was not the case in the Kirkegaard et al. (2019) or Lasker et al. (2019) studies, perhaps due to low variance in admixture. Nevertheless, the evidence is strong that studies of genetic admixture can detect a relationship between individuals' continental genetic ancestry source and cognitive test scores.

This study benefits from being pre-registered (though the final sample size was smaller than intended and subgroup analysis was not possible due to low statistical power) and the robustness of the results to different methods of data preparation. Thus, the results of the study are unlikely to be the result of questionable research practices, such as HARKing (Kerr, 1998) or *p*-hacking (John, Loewenstein & Prelac, 2012). Readers can verify my results or perform new analyses because the data and my analysis syntax are available on the OSF web site.

4.4. Alternative Interpretations

One possible alternative explanation of these results is that skin color could be a confounding variable. If American society behaves in discriminatory ways towards people with darker skin within minority groups, then this might masquerade as the genetic effect I observed in this study. This "colorism" interpretation is implausible for two reasons. The first is that heritability is highly similar for European Americans (58%), Hispanic Americans (73%), and African Americans (60%; Pesta et al., 2020, Table 3). Thus, the balance between genetic and environmental influences on phenotype variance is similar for all three groups, which makes it unlikely that a powerful discriminatory variable would operate uniquely within one group. Additionally, the cognitive ability phenotype for this study was a test score, and it is not clear how an objectively scored, computer-administered test would "know" to discriminate by giving a lower test score to a Hispanic or African American with darker skin than one with lighter skin. A cognitive ability test score is much less likely to be subjected to the buffetings of discrimination when compared to a more socially determined variable, such as income or occupational prestige. In their admixture study, Kirkegaard et al. (2019) debated the colorism hypothesis and found it insufficient to explain their similar results, citing research showing that within-family examinations of the colorism hypothesis show very weak or null effects for skin color influences. But the most convincing evidence of all that skin color is not a confounding variable comes from Lasker et al. (2019), who controlled for skin color in their admixture study and found that it did not attenuate the correlation between European ancestry and cognitive ability scores.

Another possible non-genetic explanation for my study's results is that the correlations could be caused by culture, which would be plausible if sample members with higher levels of European ancestry were more acculturated into European/Western culture than people with lower levels of European ancestry. While the data in this study do not eliminate this possibility, the Kirkegaard et al. (2019) and Lasker et al. (2019) studies undermine a culture-only interpretation of these results because controlling for self-identified race/ethnicity — a proxy for identification with a minority culture — in those studies did not explain the relationship between ancestry proportions and cognitive ability. Regardless of covariates included in Kirkegaard et al.'s (2019) and Lasker et al.'s (2019) statistical models, relationships with cognitive ability were consistently positive for European ancestry and consistently negative for African and Native American ancestry. Another complication to interpreting the correlations in this study as being the product of culture is that culture is, partially, a byproduct of evolution (Cochran & Harpending, 2009; Rindermann, 2018; Warne, 2019). If some of these cultural practices have their roots in the evolutionary history of sample members' ancestral populations, then even these cultural differences may have an ultimate genetic cause. This is one of the reasons why within-group admixture studies are so important: they reduce cultural differences because sample members belong to the same subculture within a diverse nation and experience similar treatment from society and outsiders.

One frequent objection to results of this kind is that the observed correlations have no causal implications, being an artifact of the long-term geographic separation of the sample's parent populations. The alleles used to identify genetic continental ancestry are unlikely to have any causal effects on cognitive ability. Indeed, given the realities of genetic drift, linkage disequilibrium, and the diminished predictiveness of polygenic scores derived from Europeans when used to predict phenotypes in people from other ethnic and racial groups, it is likely that virtually none of the specific alleles used by commercial companies to estimate customers' genetic ancestry have any causal impact on cognitive ability.

However, a specific causal link between these alleles and cognitive ability is not necessary for these correlations to represent a genetic influence on cognitive ability differences between groups. Both the polymorphisms used for ancestry testing and those affecting cognitive ability are dispersed widely across the genome, such that the proportion of European ancestry-informative markers in the genome is a rather close match to the proportion of European-origin cognition-related alleles. Readers who believe that the results of this study and the other recent admixture studies (Lasker et al., 2019; Kirkegaard et al., 2019) are mere artifacts with no causal connection to cognitive ability should use genomic data to demonstrate that their beliefs are empirically better supported. Until then, the straightforward interpretation — that the correlation observed in admixture studies is due to a partially genetic cause — should be the default hypothesis.

4.5. Further research . . . and a testable hypothesis

The next step in research on this topic is to continue building up admixture studies in the literature, either with exact replications of this study or with more precise or reliable sources of data. The continuing trend towards larger sample sizes will strengthen research on this topic, which will allow researchers to better control for data artifacts. Another worthwhile endeavor would be to conduct the study in a location where the population is more racially admixed, such as Brazil or Hawaii. This would provide greater variability and increase the likelihood of identifying a correlation between non-European ancestry and cognitive ability. Indeed, this study may not generalize to other nations, other racial/ethnic groups, or in a worldwide sample because of variation in environmental and genetic characteristics worldwide.

Although I found a correlation between sample members' proportion of European ancestry and their cognitive test scores, these results do not imply that there is anything special about European genetic heritage. Most alleles in Europeans (and their descendants) are found throughout the world's indigenous populations (1000 Genomes Project Consortium, 2015). Instead, I hypothesize that Europeans differ from Africans and Native Americans mostly in the *frequency* of alleles associated with high cognitive ability — on average — in their genome. This would explain why polygenic scores derived from genome-wide association studies (GWAS) on Europeans can still predict educational attainment or cognitive ability in non-European samples, though with reduced accuracy (Domingue et al., 2015; Piffer, 2015).

This hypothesis has two testable implications. First, any person with a high proportion of such alleles would also have a higher probability of scoring above average on cognitive tests, regardless of the individual's geographic ancestry. This implication can be tested with GWAS research on samples consisting of non-European racial or ethnic groups to identify alleles associated with higher cognitive ability across groups. Polygenic scores derived from diverse samples would provide stronger evidence for my hypothesis because alleles that are common in other populations may be undersampled in European groups. Sampling these populations better in the derivation of polygenic scores would provide a robust test of my hypothesis that Europeans merely differ from Native Americans and Africans in the frequency of alleles associated with higher cognitive ability.

A second test of my hypothesis arises from the fact that both East Asians and Ashkenazi European Jews have higher mean cognitive ability scores than other Europeans (Carman, Walther & Bartsch, 2018; Curran & Kellogg, 2016; Goodenough, 1926; Kane, 2007; Lynn, 2015; Lynn & Cheng, 2013; Olszewski-Kubilius & Lee, 2011; te Nijenhuis et al., 2014). Some theorists have proposed that genetic differences across these groups may partially account for these mean group differences (e.g., Cochran et al., 2006; Cochran & Harpending, 2009; Rushton, 2000; Unz, 2003). Therefore, if group differences in cognitive ability partially arise from differing frequencies of alleles associated with high ability, then East Asians and European Jews should have — on average — higher frequencies of these alleles than other Europeans. Indeed, a newly published study comparing polygenic scores for cognitive ability in European Jews with other Europeans shows exactly this result (Dunkel et al., 2019; Piffer 2019, Table 5). Similarly, within the subgroup of Europeans who also have East Asian or European Jewish ancestral heritage, the proportion of East Asian or European Jewish heritage is expected to correlate positively with cognitive ability test scores.

If these hypotheses are supported in later admixture studies, then a deeper dive into genomic data in diverse samples would be a fruitful avenue of research. With full genome sequencing of diverse samples (including non-admixed individuals), it would be possible to ascertain regions of the genome that are associated with higher cognitive ability in different populations. This detailed level of data would also be necessary to separate alleles that have a causal relationship with cognitive ability from those that are associated with causal alleles through linkage disequilibrium. Finally, understanding the impact of rare genetic variants may prove to be a fascinating contribution to the study of between-group sources of variance in cognitive ability. If any causal alleles are pleiotropic (which is likely), then this may provide clues to the evolutionary pressures that resulted in higher average cognitive ability for some human populations.

4.6. Conclusion

For the past four decades, circumstantial evidence has mounted that between-group differences in general cognitive ability are at least partially genetically influenced (e.g., Dalliard, 2014; Jensen, 1985, 1998; Lubke et al., 2003; Rowe & Cleveland, 1996; Warne, in press). This study adds to that evidence, with one important new component: the independent variable in this study is a direct measure of individuals' ancestry. By definition, this variable is genetic and therefore provides strong evidence that intergroup differences in general mental ability are partially genetically caused. With the results of this study already being replicated (Kirkegaard et al., 2019; Lasker et al., 2019), the major question now is the magnitude of these influences, which — like any heritability estimate — probably depends somewhat on the groups being compared and the geographic location in which the study occurs.

Acknowledgement: Funding for this study was provided through a Utah Valley University Grant for Engaged Learning.

Pre-registration

Study pre-registration and study materials, except the confidential cognitive test items, are available at https://osf.io/bc5fv/

References

1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526: 68-74.

Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K.E., Nivard, M.G., Veul, L., . . . & Visscher, P.M. (2019). Genetic correlates of social stratification in Great Britain. *Nature Human Behavior* 3: 1332-1342.

Adhikari, K., Chacón-Duque, J.C., Mendoza-Revilla, J., Fuentes-Guajardo, M. & Ruiz-Linares, A. (2017). The genetic diversity of the Americas. *Annual Review of Genomics and Human Genetics* 18: 277-296.

Bouchard, T. Jr. (2014). Genes, evolution and intelligence. *Behavior Genetics* 44: 549-577.

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics* 96: 37-53.

Carman, C.A., Walther, C.A.P. & Bartsch, R.A. (2018). Using the Cognitive Abilities Test (CogAT) 7 nonverbal battery to identify the gifted/talented: An investigation of demographic effects and norming plans. *Gifted Child Quarterly* 62: 193-209.

Cochran, G., Hardy, J. & Harpending, H. (2006). Natural history of Ashkenazi intelligence. *Journal of Biosocial Science* 38: 659-693.

Cochran, G. & Harpending, H. (2009). *The 10,000 Year Explosion: How Civilization Accelerated Human Evolution*. New York, NY: Basic Books.

Condon, D.M. & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence* 43: 52-64.

Curran, F.C. & Kellogg, A.T. (2016). Understanding science achievement gaps by race/ethnicity and gender in kindergarten and first grade. *Educational Researcher* 45: 273-282.

Dahlke, J.A. & Sackett, P.R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology* 102: 1403-1420.

Dalliard, M. (2014). The elusive x-factor: A critique of J. M. Kaplan's model of race and IQ. *Open Differential Psychology*. Retrieved from https://openpsych.net/files/papers/Dalliard_2014a.pdf

Domingue, B.W., Belsky, D.W., Conley, D., Harris, K.M. & Boardman, J.D. (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA Open* 1(3): 1-13.

Dunkel, C.S., Woodley of Menie, M.A., Pallesen, J. & Kirkegaard, E.O.W. (2019). Polygenic scores mediate the Jewish phenotypic advantage in educational attainment and cognitive ability compared with Catholics and Lutherans. *Evolutionary Behavioral Sciences* 13: 366-375.

Fuerst, J. & Kirkegaard, E.O.W. (2016). Admixture in the Americas: Regional and national differences. *Mankind Quarterly* 56: 255-373.

Goodenough, F.L. (1926). Racial differences in the intelligence of school children. *Journal of Experimental Psychology* 9: 388-397.

Gottfredson, L.S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law* 11: 311-319.

Han, E., Carbonetto, P., Curtis, R.E., Wang, Y., Granka, J.M., Byrnes, J., … & Ball, C.A. (2017). Clustering of 770,000 genomes reveals post-colonial population structure of North America. *Nature Communications* 8: 14238.

Herrnstein, R.J. (1971). I.Q. *The Atlantic* 228(3): 43-64.

Herrnstein, R.J. & Murray, C. (1996). *The Bell Curve: Intelligence and Class Structure in American Life*, 2nd ed. New York, NY: Free Press.

International Cognitive Ability Resource Team (2014). *The International Cognitive Ability Resource*. https://icar-project.com

Jacobs, N., Gestel, S.V., Derom, C., Thiery, E., Vernon, P., Derom, R. & Vlietinck, R. (2001). Heritability estimates of intelligence in twins: Effect of chorion type. *Behavior Genetics* 31: 209-217.

Jensen, A.R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences* 8: 193-219.

Jensen, A.R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.

Jin, W., Xu, S., Wang, H., Yu, Y., Shen, Y., Wu, B. & Jin, L. (2012). Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Research* 22: 519-527.

John, L.K., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science* 23: 524-532.

Kane, H. (2007). Group differences in nonverbal intelligence: Support for the influence of Spearman's *g*. *Mankind Quarterly* 48: 65-82.

Kerr, N.L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review* 2: 196-217.

Kirkegaard, E.O.W. & Fuerst, J. (2016). Inequality in the United States: Ethnicity, racial admixture and environmental causes. *Mankind Quarterly* 56: 580-606.

Kirkegaard, E.O.W. & Fuerst, J. (2017). Admixture in Argentina. *Mankind Quarterly* 58: 542-580.

Kirkegaard, E.O.W., Woodley of Menie, M.A., Williams, R.L., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych* 1(1): 1-25.

Lasker, J., Pesta, B.J., Fuerst, J.G. & Kirkegaard, E.O.W. (2019). Global ancestry and cognitive ability. *Psych* 1(1): 431-459.

Lee, J.J. (2010). Review of *Intelligence and How to Get It: Why Schools and Cultures Count* by R.E. Nisbett. *Intelligence* 48: 247-255.

León, F.R. & Burga-León, A. (2015). How geography influences complex cognitive ability. *Intelligence* 50: 221-227.

Lewontin, R.C. (1970). Race and intelligence. *Bulletin of the Atomic Scientists* 26: 2-8.

Lima-Costa, M.F., Rodrigues, L.C., Barreto, M.L., Gouveia, M., Horta, B.L., Mambrini, J., … & Tarazona-Santos, E. (2015). Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). *Scientific Reports* 5: 9812.

Loehlin, J.C., Vandenberg, S.G. & Osborne, R.T. (1973). Blood group genes and Negro-White ability differences. *Behavior Genetics* 3: 263-270.

Lubke, G.H., Dolan, C.V., Kelderman, H. & Mellenbergh, G.J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence* 31: 543-566.

Lynn, R. (2015). *Race Differences in Intelligence: An Evolutionary Analysis*, 2nd ed. Arlington, VA: Washington Summit Publishers.

Lynn, R. & Cheng, H. (2013). Recent data for majority and racial minority differences in intelligence of 5-year olds in the United Kingdom. *Intelligence* 41: 452-455.

Mersha, T.B. & Abebe, T. (2015). Self-reported race/ethnicity in the age of genomic research: Its potential impact on understanding health disparities. *Human Genomics* 9(1): 1-15.

Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., … & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist* 51: 77-101.

Nisbett, R.E. (2009). *Intelligence and How to Get It: Why Schools and Cultures Count*. New York: Norton & Co.

Nisbett, R.E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D.F. & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist* 67: 130-159.

Novembre, J. & Peter, B.M. (2016). Recent advances in the study of fine-scale population structure in humans. *Current Opinion in Genetics & Development* 41: 98-105.

Olszewski-Kubilius, P. & Lee, S.-Y. (2011). Gender and other group differences in performance on off-level tests: Changes in the 21st century. *Gifted Child Quarterly* 55: 54-73.

Pesta, B.J., Kirkegaard, E.O.W., te Nijenhuis, J., Lasker, J. & Fuerst, J.G.R. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence* 78: 101408.

Piffer, D. (2013). Correlation of the COMT Val158Met polymorphism with latitude and a hunter-gather lifestyle suggests culture–gene coevolution and selective pressure on cognition genes due to climate. *Anthropological Science*: 130731.

Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence* 53: 43-50.

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from Gwas hits: A replication of previous findings using recent data. *Psych* 1(1): 55-75.

Plomin, R., DeFries, J.C., Knopik, V.S. & Neiderhiser, J.M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science* 11: 3-23.

Rindermann, H. (2018). *Cognitive Capitalism: Human Capital and the Wellbeing of Nations*. New York, NY: Cambridge University Press.

Rindermann, H., Becker, D. & Coyle, T. (2016). Survey of expert opinion on intelligence: Causes of international differences in cognitive ability tests. *Frontiers in Psychology* 7: 399.

Rindermann, H., Becker, D. & Coyle, T.R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence* 78: 101406.

Roth, P.L., Bevier, C.A., Bobko, P., Switzer, F.S. III & Tyler, P. (2001). Ethnic group difference in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology* 54: 297-330.

Rothhammer, F. & Llop, E. (1976). Amerindian descent and intellectual performance in Chilean university students. *Human Biology* 48: 455-464.

Rowe, D.C. & Cleveland, H.H. (1996). Academic achievement in Blacks and Whites: Are the developmental processes similar? *Intelligence* 23: 205-228.

Rushton, J.P. (2000). *Race, Evolution, and Behavior: A Life History Perspective*, 3rd ed. Port Huron, MI: Charles Darwin Research Institute.

Rushton, J.P. & Jensen, A.R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law* 11: 235-294.

Scarr, S., Pakstis, A.J., Katz, S.H. & Barker, W.B. (1977). Absence of a relationship between degree of white ancestry and intellectual skills within a black population. *Human Genetics* 39: 69-86.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science* 174: 1285-1295.

Shuey, A.M. (1966). *The Testing of Negro Intelligence*, 2nd ed. New York, NY: Social Science Press.

Smouse, P.E., Spielman, R.S. & Park, M.H. (1982). Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. *American Naturalist* 119: 445-463.

Snyderman, M. & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist* 42: 137-144.

Tang, H., Quertermous, T., Rodriguez, B., Kardia, S.L.R., Zhu, X., Brown, A., … & Risch, N.J. (2005). Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *American Journal of Human Genetics* 76: 268-275.

te Nijenhuis, J., David, H., Metzen, D. & Armstrong, E.L. (2014). Spearman's hypothesis tested on European Jews vs non-Jewish Whites and vs Oriental Jews: Two meta-analyses. *Intelligence* 44: 15-18.

te Nijenhuis, J. & van den Hoek, M. (2016). Spearman's hypothesis tested on black adults: A meta-analysis. *Journal of Intelligence* 4(2): 6.

Thoday, J.M. (1969). Limitations to genetic comparison of populations. *Journal of Biosocial Science* 1(S1): 3-14.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., … & Williams, S.M. (2009). The genetic structure and history of Africans and African Americans. *Science* 324: 1035-1044.

Unz, R. (2013, March/April). How social Darwinism made modern China. *The American Conservative*. https://www.theamericanconservative.com/articles/how-social-darwinism-made-modern-china-248/

Warne, R.T. (2016). Testing Spearman's hypothesis with Advanced Placement examination data. *Intelligence* 57: 87-95.

Warne, R.T. (2019). Review of *Cognitive Capitalism: Human Capital and the Wellbeing of Nations*, by H. Rindermann. *Intelligence* 73: 63-64.

Warne, R.T. (in press). *In the Know: Debunking 35 Myths about Human Intelligence*. New York, NY: Cambridge University Press.

Winegard, B., Winegard, B. & Boutwell, B. (2017). Human biological and psychological diversity. *Evolutionary Psychological Science* 3: 159-180.

Witty, P.A. & Jenkins, M.D. (1936). Intra-race testing and negro intelligence. *Journal of Psychology* 1: 179.

Woodley, M.A., Rindermann, H., Bell, E., Stratford, J. & Piffer, D. (2014). The relationship between *Microcephalin*, *ASPM* and intelligence: A reconsideration. *Intelligence* 44: 51-63.

Woodley of Menie, M.A., Younuskunju, S., Balan, B. & Piffer, D. (2017). Holocene selection for variants associated with general cognitive ability: Comparing ancient and modern genomes. *Twin Research and Human Genetics* 20: 271-280.

CHAPTER 5

BIOGEOGRAPHIC ANCESTRY, COGNITIVE ABILITY AND SOCIOECONOMIC OUTCOMES[1]

EMIL O. W. KIRKEGAARD, ROBERT L. WILLIAMS,
JOHN G.R. FUERST AND GERHARD MEISENBERG

Abstract

The cause(s) of ubiquitous cognitive differences between American self-identified racial/ethnic groups (SIREs) is uncertain. Evolutionary-genetic models posit that ancestral genetic selection pressures are the ultimate source of these differences. Conversely, sociological models posit that these differences result from racial discrimination. To examine predictions based on these models, we conducted a global admixture analysis using data from the Pediatric Imaging, Neurocognition, and Genetics Study (PING; N = 1,369 American children). Specifically, we employed a standard methodology of genetic epidemiology to determine whether genetic ancestry significantly predicts cognitive ability, independent of SIRE. In regression models using four different codings for SIRE as a covariate, we found incremental relationships between genetic ancestry and both general cognitive ability and parental socioeconomic status (SES). The relationships between global ancestry and cognitive ability were partially attenuated when parental SES was added as a predictor and when cognitive ability was the outcome. Moreover, these associations generally held when subgroups were analyzed separately. Our results are congruent with evolutionary-genetic models of group differences and with certain environmental models that mimic the predictions of evolutionary-genetic ones. Implications for research on race/ethnic differences in the Americas are discussed, as are methods for further exploring the matter.

Keywords: race; ethnicity; biogeographic ancestry; genomics; cognitive ability; socioeconomic status

1. Introduction

Cognitive ability, whether measured by IQ tests, Piagetian tests, educational/scholastic tests (e.g., PISA, TIMSS, etc.), or other indices of cognitive functioning, differs on average between biogeographic ancestry groups (BGAs). (We use the term "cognitive ability" instead of "general intelligence" or "general cognitive ability" to make clear that we are not committing to a claim about the psychometric nature of the underlying construct. For the purpose of this study, the distinctions between cognitive ability in general and general cognitive ability are not important.) According to Shriver and Kittles (2004), biogeographic ancestry is the personal genetic history, indexed by ancestrally informative autosomal markers, which reflects an individual's overall ancestry with respect to "population groups" (also referred to as "clusters," "ancestral groups," and "ancestral populations”). These reference biogeographic ancestry groups differ owing to the effects of evolutionary factors, such as “isolation by distance,” and barriers that “have all affected human migration and mating patterns in the past.” Biogeographic ancestry groups have commonly been called “races” (see, Fuerst, 2017; Woodley, 2010). Examples of these groups include East Asians, Europeans, sub-Saharan Africans, and Amerindians (see, Christainsen, 2013). Cognitive ability differences appear both between nations (e.g., in predominantly European versus predominantly African national populations), and between self-identified racial and ethnic (SIRE) groups within nations (e.g., Whites versus African Americans in the USA (Fuerst & Kirkegaard, 2016a, 2016b; Lynn, 2015). The cause(s) of inter- and intra-national cognitive ability gaps is highly uncertain and hotly debated, with some researchers arguing that genetic differences are substantially or even predominantly responsible for at least some of these gaps (e.g., Rushton & Jensen, 2005) and others maintaining that non-genetic environmental factors fully account for the gaps (e.g., Cottrell et al., 2015; Nisbett et al., 2012).

---

[1] Editors' note: We have reformatted the original paper for this collection, updated the references to APA style, and made other changes as requested by the authors of the original paper.

Nevertheless, in the most recent survey of intelligence researchers, “genes were rated as the most important cause (17%) [of international differences], followed by educational quality (11.44%), health (10.88%), and educational quantity (10.20%)” (Rindermann et al., 2016). Further, 90% of the surveyed experts believed that international differences were at least in part genetic in origin. Rindermann, Coyle, and Becker (2013) also reported that 83% of survey respondents believed that the Black-White cognitive gap in the USA is partially due to genetic differences between the relevant populations. While there certainly is no general consensus among researchers, a substantial number of them believe that both international and certain intra-national SIRE differences in cognitive ability have a partial genetic basis (Rindermann et al., 2016; Snyderman & Rothman, 1988)

Although racial and ethnic groups within a country can differ behaviorally for a variety of reasons (e.g., selective self-identification or local selective migration), possible genetic bases of these differences are typically analyzed in an evolutionary framework. When discussing evolved human diversity, one can conceptualize groups in a number of different ways, e.g., as subspecies (taxonomically significant subdivisions of a species), ecotypes (environmentally adapted types), clines (character gradients), and morphs (alternative phenotypes in a population). Following the work of Charles Darwin, communities are frequently delineated by propinquity descent since descent is understood as inductively potent. The preferable term to describe descent-based groups (e.g., variety, genetic population, race, genetic cluster, ancestry group, etc.) is a matter of ongoing semantic dispute. Here we call them “biogeographic ancestry groups” (again, BGAs), as is frequently done in genetic epidemiology. There are phenotypic and genetic differences between these groups, with evolutionary forces acting over relatively short spans of time likely having caused this divergence (Cochran & Harpending, 2009). Importantly, these ancestry groups are delimitable using genomic data.

The cause(s) of inter-BGA differences in cognitive ability specifically is the subject of several evolutionary theories, which decompose broadly into two categories: (1) Pleistocene-selection models and (2) Holocene-selection models. The former posit the action of evolutionarily novel factors that differed systematically in the regions to which human groups migrated after leaving Africa around 60 to 100 kya (thousand years ago), and which specifically selected for increased cognitive ability as a mechanism for enhancing survival (Kanazawa, 2012). Salient evolutionarily novel challenges likely included the presence of seasonality, specifically cold winters during the main Würm glaciation event 60 kya, when temperatures in Europe and (especially) Northeast Asia were considerably lower on average than they are today. Extreme cold, coupled with the challenge of provisioning for the future such as to anticipate seasonality (e.g., food storage) have been proposed as sources of selection for higher cognitive ability and other somatic traits related to innovation and productivity (Grall, 2016; Wallace, 1864). These climatic models predict that North East Asian and European BGA groups will have higher levels of cognitive ability than Pacific Islanders, sub-Saharan Africans, and Amerindians (at least from the tropical and subtropical regions of the Americas).

Research in comparative zoology supports climatic models of this kind. A large body of research has investigated the climatic correlates of cognitive ability—usually measured with brain size as a proxy variable—in non-human animals. Non-migratory birds have been a primary choice of study (Sayol et al., 2016; Sol et al., 2005, 2007, 2010) given their range of habitat and the ease of studying them. Studies have revealed the expected patterns, namely that birds that live further north and in more seasonally affected areas have larger brains (controlling for body size), more flexible behavior, and more innovative behavior (Roth & Pravosudov, 2009; Sayol et al., 2016; Sol et al., 2005).

Holocene-selection models stress the observation that the rate of adaptive evolution during the Holocene (beginning roughly 12 kya) was on the order of 100 times that experienced by humans during the preceding Pleistocene epoch, thus the evolutionary factors crucial to the origin of these differences may have arisen relatively recently (Cochran & Harpending, 2009). Culture-gene co-evolution theory posits that cognitive ability and cultural complexity arose in tandem via a feedback process. Eurasian populations' transition into sedentarism during the late Pleistocene and early Holocene would have been associated with a substantial increase in cognitive challenges, such as those related to competition over settled land, management of agriculture, and sustainment of higher standing population densities (Cochran & Harpending, 2009; Rindermann & Woodley, 2012). Innovations in modes of production and social organization (such as intensified division of labor) would have led to hierarchical societies involving strong individual-level competition for finite resources, which may have been a major source of social selection favoring higher levels of cognitive ability. Clark (2007, 2014) documents the persistence of downward social mobility among the descendants of elites who competed for limited economic and social resources, with the less competitive sinking into lower-status occupational niches. The end result of this process was a persistent "bootstrapping" of the population, as cognitive ability (and other salient "bourgeois" traits) rose across all levels of the social hierarchy, leading to greater degrees of industriousness and innovativeness. This may in turn have boosted the competitiveness of certain European, and perhaps Northeast Asian, BGAs, permitting group-level expansion (or colonialism), which likely enhanced the biocultural "corporate" fitness of these populations—a process that may only have ceased in the West in the mid-19th century. Around this time, the advent of milder climates and concomitantly reduced ecological stress (corresponding to the end of the Maunder Minimum), coupled with both social and technological innovations that allowed those with relatively high cognitive ability to control their fertility, potentially reversed the fitness advantage of those with relatively high cognitive ability compared to those with relatively low cognitive ability; Woodley et al., 2017). Given that these Holocene-selection models posit that Eurasian populations had the greatest exposure to environmental challenges favoring the fitness of those with relatively high cognitive ability, they predict that selection for cognitive ability was stronger in Eurasian populations than non-Eurasian ones, such as Pacific Islanders, sub-Saharan Africans, and Amerindians.

A further evolutionary theory of inter-BGA cognitive differences is the disease-stress model. This has assumed forms that posit (Fedderke et al., 2014) and do not posit (Eppig et al., 2010) genetic differences to explain inter-BGA disparities in cognitive ability. Models of the latter type propose that in the Pleistocene environment of evolutionary adaptedness, humans were repeatedly exposed to periods of high and low parasite stress. They further maintain that this exposure to variable levels of parasite stress favored the evolution of epigenetic adaptations that dynamically regulate tradeoffs of bioenergetic investments into brain/cognitive development or immune system functioning in response to environmental challenges, with greater investments into the latter over the former occurring in high-parasite-stress contexts. Therefore, when humans radiated into novel, more northerly and easterly low-parasite-stress ecologies, they came pre-adapted with the capacity to developmentally trade immune system functioning for higher cognitive ability, which is reflected in very strong inverse ecological correlations between national IQ and parasite-burden indices (Eppig et al., 2010). Given the global distribution of parasite burdens, this model predicts, as do the other evolutionary theories, that the populations of Europe and North East Asia have higher cognitive ability relative to those of the Pacific Islands and the tropical and subtropical regions of the Americas, as well as the populations of sub-Saharan Africa. But this version of the disease-stress model is clearly inadequate to account for the total global pattern of inter-BGA cognitive ability variation, since it seemingly cannot explain BGA cognitive ability gaps found within nations, where BGA groups do not differ substantially in their exposure to parasites (since they inhabit the same general environments that the nations encompass). Therefore, our statistical analyses do not endeavor to test predictions of this disease-stress model.

Conversely, genetic disease-stress models posit that globally variable geographical factors covary with disease burdens, and that genetic adaptations to local disease burdens have partially generated inter-BGA differences in social outcomes and related phenotypes, including cognitive ability (Fedderke et al., 2014). While our data do not permit any direct test of this, or any other, particular evolutionary account of the emergence of inter-BGA cognitive ability variation, global admixture analytic findings indicative of a genetic etiology of such variability would be consistent with all the evolutionary theories reviewed here, apart from the epigenetic disease-stress model.

1.1. Genomic Studies of Selection in Humans

There are several fairly well-documented cases of recent simple selection in humans. Examples include skin tone and high-altitude adaptations. The study of highly polygenic selection via changes in frequencies of many alleles (soft sweeps) in humans has advanced recently due to the availability of large-scale, powerful computer clusters and large genomic datasets. A number of recent studies based on genomic data have found evidence of recent polygenic selection in humans over time (Mathieson et al., 2015; Woodley et al., 2017) and space (Piffer, 2013, 2015; Robinson et al., 2015; Turchin et al., 2012). These findings indicate selection for, e.g., height (Turchin et al., 2012), body mass (Robinson et al., 2015), cognitive ability (Piffer, 2015), educational attainment (Piffer, 2013; Racimo et al., 2018), and schizophrenia (Guo et al., 2018).

One significant concern with genomic studies that compare polygenic score (PGS) frequencies between major ancestry groups is the "transethnic validity" of PGS (Ntzani et al., 2012). Most genome-wide association studies (GWASs) are done solely on people of European ancestry (to avoid population-stratification confounding). These studies unfortunately yield predictive models that are less useful in non-Europeans, particularly Africans (Marigorta & Navarro, 2013; Zanetti & Weale, 2016). This problem results from two factors. First, the genetic variants discovered in GWASs are typically not causal (with respect to the trait of interest) but are so-called "tag variants." These are variants that are close enough on the genome to the causal variants such that their presence in the population is statistically linked (known as linkage disequilibrium; LD). Second, the degree to which two variants on the genome are in LD partly depends on the populations studied and their genetic distance; this is because random genetic patterns can arise over time (i.e., genetic drift in LD).

The effect is that a given variant which tags a causal variant in Europeans may not do so, or not very well, in other populations, thus reducing the validity of European-derived estimates of the true PGS. As a result, the predictive validity of associated variants in one major BGA group frequently does not transfer to others. While a number of methods have been employed to control for drift-related effects (Guo et al., 2018; Piffer, 2015), and while robust or partially robust inter-BGA differences in educational attainment and intelligence PGS have been found, until causal variants that are not affected by LD (Marigorta & Navarro, 2013) are identified, PGS-based results will continue to carry a degree of uncertainty.

1.2. Admixture Analysis

Admixture analysis—analysis of genetic ancestry in previously isolated but recently interbred populations, which relates genetic ancestry to outcomes—is a potent tool used by medical geneticists for the exploration of the source of trait differences and disease disparities in and among admixed populations. The relationships between genomic ancestry and phenotype are treated as indirect evidence of genetic causation, especially when confounds are controlled in regression analysis (e.g., Bidulescu et al., 2014; Giri et al., 2017; Meigs et al., 2014). Admixture analysis includes global admixture analysis and admixture mapping. When ancestral BGA groups vary in the frequency of genetic variants underlying a trait, in admixed populations the phenotype of interest will be correlated with BGA in genomic regions near the causal genetic variants. This situation allows for the identification of associated loci, a process called admixture mapping (Shriner, 2013; Winkler et al., 2010). When the trait has a complex genetic architecture, one where thousands of loci are assumed to contribute to the phenotype, an appropriate first step is global admixture analysis. This process seeks to identify associations between global BGA and phenotype, without attempting to identify local regions of a genome associated with a phenotype. One advantage to global admixture analysis is that it requires much smaller sample sizes compared to admixture mapping.

Templeton (2001) has intricately detailed the logic of global admixture analysis as applied to evolutionary-genetic models. Generally, global admixture analysis within ethnic groups can be viewed as a Mendelian "common garden" experiment, since members of the same ethnic groups experience similar cultures and environments. Factors affecting SIRE groups in general (e.g., stereotype threat, race-based discrimination, segregation, cultural norms, dialect, language, etc.) are either controlled for or attenuated. This basic logic has been adopted by genetic epidemiologists e.g., (Bidulescu et al., 2014; Giri et al., 2017; González et al., 2005; Meigs et al., 2014) who frequently examine the association between global genetic ancestry and traits either within ethnic groups or after controlling for SIRE. Admixture analysis has been applied to study the etiology of inter-BGA differences in, among other traits, height (Becker et al., 2011), sleep behaviors (Halder et al., 2015), and brain morphology (Fan et al., 2015). The quantitative predictions and theoretical basis of global admixture analysis have often not been explicated. Doing so requires employment of a series of simulations that incrementally increase in model complexity and realism. We do this in Appendix 1 of Supplementary File 1, which contains the simulations and explanations.

Before the advent of genomic ancestry testing, attempts were made to test evolutionary predictions with respect to inter-BGA cognitive ability differences using other methods (for reviews, see Loehlin et al., 1975). Ancestry was estimated using traits such as skin color, blood groups, and reported genealogy. In other words, these studies relied on various proxies or poor measures of genomic ancestry, and further the samples were usually small; moreover, no one has systematically meta-analyzed the data to settle the interpretive dispute. More recent studies have utilized nation-level data on the frequencies of Y-chromosomal haplogroups (Rindermann & Woodley, 2012; Rodriguez-Arana, 2010), estimates of between-country genetic distance (Becker & Rindermann, 2016), and estimates of regional and national-level ancestry percentages (Fuerst & Kirkegaard, 2016a) as predictors of national variation in cognitive ability. These recent studies have been restricted to the national level of analysis, which necessarily reduces sample resolution. As a result of these problems and others, there has been disagreement about whether past studies employing these techniques show, in admixed populations, the expected relations between BGA and cognitive ability (Nisbett, 2009; Rushton & Jensen, 2005; Sternberg, 2005).

Thus far, no published study has examined genomic ancestry, assessed using genomic era methods, in a large sample and related the findings to cognitive ability, while taking into account the confounding effect of racial/ethnic identity. One study (Akshoomoff et al., 2014), using the same sample presently examined, reported ANOVA results for genetic ancestry and test scores. The authors, however, did not report coefficients for specific ancestries, did not use general factor scores, and did not examine the effect of ancestry within SIRE groups. The last of these is a critical aspect of the so-called "common garden" experiment. A large meta-analysis of pan-American epidemiological studies found that European genetic ancestry was robustly associated with better socioeconomic outcomes relative to African and Amerindian ancestry in admixed populations (European: $r = 0.18$, $k = 28$, $N = 35{,}476.5$; Kirkegaard et al., 2017). Amerindian and African ancestry were related to poorer socioeconomic outcomes: $r = -0.14$, $k = 31$, $N = 28{,}937.5$ and $r = -0.11$, $k = 28$, $N = 32{,}710.5$, respectively. Consistent with evolutionary models, these associations were found within admixed ethnic groups, and when the effect of SIRE was statistically controlled for. Given these associations between genetic ancestry and SES, and the moderate-to-strong relationship between cognitive ability and SES (Sirin, 2005), it is likely that BGA is also associated with higher cognitive ability at least partially independent of SIRE. A related finding is that BGA is a strong predictor of regional cognitive and general socioeconomic outcomes across the Americas (Fuerst & Kirkegaard, 2016a, 2016b, Kirkegaard & Fuerst, 2017) both at the country- and first-order division (e.g., province, state, district) level. This provides ecological support for the expectation that individual BGA will be robustly associated with cognitive ability.

The purpose of the current study is to test the prediction that BGA is associated with cognitive ability after statistically controlling SIRE (which indexes a number of factors commonly invoked to explain cognitive gaps between BGAs, such as racial discrimination), and to quantify the magnitudes of the individual-level associations. More generally, we sought to apply global admixture analysis, via genomic data, on American samples. We aimed to determine if associations between genetic ancestry and IQ are more consistent with commonly proposed evolutionary-genetic or social environment-based explanations for mean cognitive ability differences observed between European and East Asian descent groups relative to African, Amerindian, and Oceanian ones. Based on convergent results from pre-genomic-era studies, national-level analyses of ancestral markers, genomic PGS research, individual-level SES-admixture results, and regional-level SES/cognitive ability-admixture results, we expect that results similar to those found for SES-admixture will be found for cognitive ability.

2. Material and Methods

Data used in this study were obtained from the Pediatric Imaging, Neurocognition, and Genetics (PING) database (http://ping.chd.ucsd.edu/). The primary aim of PING was to create a well-standardized and carefully organized resource of magnetic resonance imaging (MRI) data, comprehensive genotyping data, and developmental and neuropsychological assessments for a large cohort of developing children. The PING data were based on a large sample (*N* = 1,391 with genetic ancestry data) of healthy American children of diverse ancestries (Akshoomoff et al., 2014). Participants were not nationally representative, but were rather recruited from the greater metropolitan areas of Baltimore, Boston, Honolulu, Los Angeles, New Haven, New York, Sacramento, and San Diego. Individuals who had medical conditions that could affect their development were excluded from the recruitment process.

2.1. Cognitive Ability

Participants were given the National Institute of Health toolbox cognitive tests. These tests have previously been validated (Weintraub et al., 2014). The cognitive sub-domain assessed by each test, along with the subtests' psychometric characteristics, are detailed by Akshoomoff et al. (2014). The seven cognitive tests were as follows: Dimensional Change Card Sort Test (Card Sort); Flanker Inhibitory Control and Attention Test (Flanker); Picture Sequence Memory Test (Picture Sequence); Pattern Comparison Processing Speed Test (Pattern Recognition); Oral Reading Recognition Test (Pattern Recognition); List Sorting Working Memory Test (List Sort); and Picture Vocabulary Test (Vocabulary).

Some data were missing (4% of cells, 14% of cases, had at least 1 missing data point). We imputed values for cases that were missing less than half of their data points rounded down (three or fewer). Imputation was performed using IRMI (Templ et al., 2011, 2015). After this, there were 1,369 cases with complete cognitive data and genetic ancestry (with 1,370 cases for some subtests).

Next, the effects of age was regressed out by first fitting a local regression model (LOESS; see James et al., 2013) with age as the predictor and then saving the residuals. LOESS was used for age because this approach is able to capture complex non-linear patterns in the data. For this regression, we used age at neuropsychological testing as the primary variable and we filled in missing data (*N* = 3) with age at MRI (total = 1,369 cases with *g* scores). The mean age was 11.75 (*SD* = 4.88; range = 3–21). Supplementary Figure S1 shows the distribution.

These residuals were then residualized again for sex using a standard linear model (OLS). As shown in Table 1, the age/sex corrected scores exhibited a positive manifold. The factor loadings of the subtests are shown in Table 1.

Table 1. The Pediatric Imaging, Neurocognition, and Genetics (PING) subtest factor loadings and pairwise intercorrelations (with age and sex controls)

	g **Factor Loading**	**Picture Sequence**	**List Sort**	**Pattern Recognition**	**Vocabulary**	**Reading**	**Flanker**	**Card Sort**
Picture Sequence	0.37	1.00						
List Sort	0.57	0.34	1.00					
Pattern Recognition	0.38	0.13	0.18	1.00				
Vocabulary	0.57	0.22	0.34	0.13	1.00			
Reading	0.55	0.20	0.31	0.13	0.56	1.00		
Flanker	0.49	0.10	0.22	0.29	0.17	0.17	1.00	
Card Sort	0.57	0.16	0.30	0.32	0.22	0.21	0.52	1.00

Note: *N* = 1,369–1,370.

Finally, we factor analyzed (minimal residual factor analysis/PAF) the adjusted data for the seven tests and extracted a general cognitive ability (*g*) factor (Jensen, 1998), using the psych 1.6.9

package (Revelle, 2017). The pattern of factor loadings looked normal and the *g* factor explained 26 percent of the variance. Using this method, *g* scores for 1,369 individuals were derived. Scores on this factor were saved and standardized ($M = 0$, $SD = 1$) for further analysis. We were then faced with the choice of whether to standardize the scores to the White subsample or whether to use the entire sample. We chose to use the full sample because this would counterbalance the likely cognitive ability/social status selection of the sample due to urban sampling. Sample sizes and descriptive statistics for the factor analysis are provided in Tables S13–S16 of Supplementary File 2.

The intercorrelations and *g* loadings were low compared to those found with standard IQ batteries; for example, the fourteen Woodcock–Johnson-III subtests showed an average *g* loading of 0.60 (Murray, 2007), while our seven subtests showed an average of 0.50 (0.37 to 0.57). These results, though, were comparable to those found in some adult datasets, such as the Human Connectome Project, which used similar subtests (Van der Linden et al., 2017). As the *g* scores for this sample predict parental SES no less well than typical IQ scores, and as they exhibit more or less the same magnitudes of ethnic differences as found using standard IQ batteries, there is no strong reason to suspect that these scores are particularly unreliable.

It is possible that there may be measurement invariance (MI) issues between SIRE groups. We do not attempt to explore this issue here, as it is not directly relevant to the hypothesis we test. The question we investigate is whether socioeconomic and cognitive differences between SIRE groups are, to some degree, statistically explainable by genetic ancestry, not whether the differences have the same sociological and psychometric meaning as differences between individuals within SIRE groups. More generally, we conceptualize our *g* scores as statistical constructs which represent summaries of observed subtest scores. We do not imply that these score differences index differences in a causal latent general factor ("biological *g*"; Kan, 2012). The same consideration applies with respect to our general SES scores. As such, more detailed analyses regarding the relation between observed scores, latent factors, and SIRE/genetic ancestry are unnecessary for testing the hypothesis (Dolan, pers. comm., 1 January 2018). Nonetheless, to assess factorial invariance, we computed Tucker's congruence coefficients for standard SIRE groups with sample sizes greater than 100. The congruent coefficient matrix is shown in Table 2. The congruence coefficient is a measure of factor similarity, with 0.95 or greater indicating virtually identical to identical factors and 0.85 to 0.94 indicating fair factor similarity (Lorenzo-Seva & Ten Berge, 2006). For only two comparisons, Asian-African American and Asian-Hispanic, were the coefficients below 0.95. Note, these results do not directly address the issue of factor invariance with respect to genetic ancestry.

Table 2. Congruence coefficients for *g* loadings across standard self-identified racial/ethnic (SIRE) groups with *N*s > 100

	White	**African American**	**Hispanic**	**Multi-Ethnic**	**Asian**
White		0.95	0.96	0.98	0.97
African American			0.99	0.96	0.87
Hispanic				0.98	0.91
Multi-ethnic					0.97
Asian					

2.2. Parental Socioeconomic Status

The following SES variables were available: household income, guardian 1 educational level, guardian 2 educational level, guardian 1 occupational level, and guardian 2 occupational level. For 95% of the participants for which relationship data were available, guardian 1 was either a biological mother or father (missing data: 2%; self: 1%) and for 92% of the participants for which relationship data were available, guardian 2 was either a biological mother or father (missing data: 32%). Since most often guardians were parents, SES is referred to as parental SES. As before, some data were missing. To maximize the available sample size, data with at least one data point were imputed. Finally, variables were factor analyzed to extract a general SES factor. Scores (1380; 1363 with *g* scores also) were saved ($M = 0$, $SD = 1$) for further analysis. As before, the standardization was done the full sample, rather than the White subset. Intercorrelations and descriptive statistics for the factor analysis are provided in Tables S17 and S18 of Supplementary File 2.

2.3. SIRE and Genetic Ancestry Percentages

Genetic ancestry percentages were available in the PING data files (see Akshoomoff et al., 2014, for details). These were calculated using over 15,000 single-nucleotide polymorphisms or SNPs (i.e., single DNA base pair variations). To access ancestry percentages, supervised clustering analysis with the ADMIXTURE algorithm was used. Ancestry was assigned to six major biogeographic clusters corresponding to indigenous Europeans, Africans, Americans, East Asians, Oceanians, and Central Asians. Ancestry percentages for a given individual summed to one with each individual being assigned a percentage with respect to each component. A total of 1391 individuals in the sample had genetic ancestry data. In addition to genetic ancestry percentages, multiple-option, dichotomous SIRE variables were available. These were based on the five major SIRE racial categories in the United States (White, African American, American Indian, Asian, and Pacific Islander) and the one US SIRE ethnic category (Hispanic). Some cases had missing SIRE data (1% of the cells); and these were imputed as “false.” A dummy variable, “Other,” was also created and set to “true” for participants who selected no SIRE. Owing to the multiple-response nature of the SIRE categories, there were a number of different methods for defining them (for discussion, see Liebler & Halpern-Manners, 2008). Since concern has been raised about the impact of a researcher’s “degree of freedom” (Wicherts et al., 2016) and about assigning people to SIRE categories (Liebler & Halpern-Manners, 2008), we coded SIRE groups in four different ways. We did this to assess how coding decisions affected outcomes. The codings were: standard coding (for which separate categories are created for Hispanic and non-Hispanic multi-racial individuals along with each of the five racial identities), common combination coding (for which a category is created for each of the unique combinations of SIRE identities), continuous or interval coding (for which individuals are assigned a percentage of each SIRE category based on the number of responses chosen), and dummy coding (for which each of the six ethnic and racial categories is used as a dichotomous categorical predictor). The methods are discussed in more detail in Supplementary File 2.

Table 3 shows the sample characteristics for the full sample (which had *g* scores) and by Standard SIRE groups. A one way ANOVA indicated a significant effect of SIRE on age at the $p < 0.01$ level for the eight SIRE categories [$F(7, 1{,}361) = 7.42$, $p < 0.01$]. A Chi-Square test also indicated a significant effect of SIRE on sex [($X^2(7, N = 1{,}369) = 21.372$, $p < 0.01$)]. As the effect of sex and age was regressed out of cognitive ability, and as sex and age have no direct effect on genomic ancestry or Parental SES, no further steps were taken.

Table 3. Sample characteristics for the full sample (with *g* scores) and by standard SIRE groups

	N	**Mean Age**	***SD* Age**	**Percent Male**	**Mean Parental SES**	***SD* Parental SES**	**Mean *g* Score**	***SD g* Score**
Full sample	1369	11.75	4.88	51.5	0.00	1.00	0.00	1.00
White	567	11.64	4.80	52.9	0.41	0.77	0.29	0.89
African American	138	12.50	4.69	55.8	−0.56	0.96	−0.68	0.88
American Indian	4	11.62	3.49	100	−1.03	0.92	0.21	1.26
Asian	120	13.81	5.67	34.2	0.20	1.04	0.27	0.88
Hispanic	323	11.38	4.85	52.9	−0.44	0.94	−0.33	0.98
Multi-ethnic	182	10.95	5.00	51.1	−0.04	1.10	0.05	1.10
Other	19	14.45	4.21	63.2	−0.15	1.00	−0.05	0.84
Pacific Islander	16	7.10	2.70	43.8	−1.24	1.20	−0.48	1.30

Note: *N* is the number of cases with *g* scores; the number additionally with socioeconomic status (SES) scores may be smaller. *SD* = standard deviation.

2.4. English as First Language

For 11% of the sample, English was not the participant's first language. As cognitive tests could be biased against non-native speakers, a dichotomous English as First Language (EFL) variable was also included as a variable.

3. Preliminary Analyses

3.1. SIRE and Genomic Ancestry

It is known that genetic ancestry varies substantially within some US SIRE groups but not others (Bryc et al., 2015). To examine the distribution of genetic ancestry by SIRE in the current dataset, means and standard deviations of genetic ancestries were calculated by standard SIRE. The results are shown in Table 4 (Table S12 of Supplementary File 2 reports more descriptive statistics for the genetic ancestries). As expected, Whites were almost entirely genomically European (97%), as has been found in previous studies of US citizens (e.g., Bryc et al., 2015). Other groups showed more admixture. For these, there were substantial discrepancies between SIRE and genetic ancestry. These discrepancies allow for the admixture analyses.

Table 4. Mean genetic ancestry by self-identified race/ethnicity (standard coding)

SIRE	*N*	**European**	**African**	**Amerindian**	**East Asian**	**Oceanian**	**Central Asian**
White	571	0.97 (0.07)	0.00 (0.04)	0.01 (0.02)	0.00 (0.04)	0.00 (0.00)	0.01 (0.04)
African American	140	0.17 (0.11)	0.81 (0.12)	0.00 (0.01)	0.00 (0.03)	0.00 (0.01)	0.01 (0.02)
American Indian	4	0.66 (0.42)	0.21 (0.42)	0.13 (0.25)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)
Asian	122	0.06 (0.17)	0.00 (0.02)	0.00 (0.01)	0.80 (0.36)	0.01 (0.01)	0.13 (0.32)
Hispanic	328	0.57 (0.23)	0.14 (0.20)	0.19 (0.17)	0.09 (0.20)	0.01 (0.02)	0.01 (0.03)
Multi-ethnic	186	0.45 (0.27)	0.11 (0.21)	0.01 (0.02)	0.37 (0.29)	0.03 (0.05)	0.03 (0.12)
Other	24	0.55 (0.36)	0.26 (0.33)	0.01 (0.03)	0.01 (0.02)	0.00 (0.01)	0.18 (0.32)
Pacific Islander	16	0.11 (0.15)	0.01 (0.04)	0.00 (0.00)	0.70 (0.13)	0.17 (0.06)	0.00 (0.00)

Note: Standard deviations in parentheses.

The correlations between continuous SIRE coding [for description, see Section 2.3 above] and genetic ancestry are shown in Table 5. It can be seen that continuous SIRE coding is a reasonably good index of genetic ancestry. Since US SIRE groups are fairly non-admixed, much of the variance is explained by recent patterns of exogamy which can be captured by this type of coding.

Table 5. Correlations between genetic ancestry and self-identified race/ethnicity (SIRE, continuous coding)

	European	African	American	East Asian	Oceanian	Central Asian
SIRE: White	**0.86**	−0.50	−0.23	−0.44	−0.22	−0.13
SIRE: African American	−0.45	**0.92**	−0.14	−0.18	−0.08	−0.06
SIRE: Hispanic	−0.08	0.00	**0.75**	−0.15	−0.06	−0.08
SIRE: Asian	−0.56	−0.20	−0.16	**0.81**	0.09	0.27
SIRE: Pacific Islander	−0.28	−0.10	−0.07	0.39	**0.81**	−0.03
SIRE: American Indian	0.00	0.03	**0.09**	−0.04	−0.01	−0.04
SIRE: Other	−0.03	0.06	−0.04	−0.06	−0.03	**0.16**

Note: Highest correlations for SIRE are in bold.

3.2. Bivariate Relationship between Genetic Ancestry, Cognitive Ability, and Parental SES

The correlations between CA, parental SES, and each of the genetic ancestries are shown in Table 6. The correlations were roughly as would be expected given the well-documented cognitive and socioeconomic differences between US SIRE groups. European ancestry had a moderate positive correlation with both cognitive ability and SES (r = 0.23 and 0.32, respectively), while negative relationships were seen for African (r = −0.33 and −0.30), Amerindian (r = −0.15 and −0.24), and Oceanian (r = −0.08 and −0.20) ancestries. The Genomic ancestry × SES correlations are larger than those previously reported by Kirkegaard et al. (2017), presumably because this sample is not decomposed by SIRE group, leading to reduced restriction of range and thus higher correlations. The remaining ancestries had very small or inconsistent relationships. Not much should be made of the bivariate analysis for ancestry and outcomes because these confound the effects of multiple ancestries, as well as the effects of non-genetic causes. In particular, the ancestry variables are necessarily negatively related in general because every individual's ancestry must sum to one (see Tables S8–S11 of Supplementary File 2 for correlation matrices by subpopulation).

Previous research has found a modest relation between parental SES and cognitive ability (Sirin, 2005). In our sample, the relation (r = 0.42) was somewhat stronger than typically reported. This increased strength of association is probably in part due to the use of a general socioeconomic factor, which is a more reliable measure, as opposed to individual measures of parental socioeconomic status.

Table 6. Pairwise correlations between cognitive ability, parental socioeconomic status (SES), and genetic ancestries

	Cognitive	**SES**	**European**	**African**	**Amerindian**	**East Asian**	**Oceanian**	**Central Asian**
Cognitive	1.00							
SES	0.42	1.00						
European	0.23	0.32	1.00					
African	−0.33	−0.30	−0.50	1.00				
Amerindian	−0.15	−0.24	−0.10	−0.08	1.00			
East Asian	0.05	−0.06	−0.62	−0.22	−0.16	1.00		
Oceanian	−0.08	−0.20	−0.31	−0.10	−0.09	0.42	1.00	
Central Asian	0.06	0.12	−0.20	−0.07	−0.07	−0.07	−0.03	1.00

Note: Sample sizes for Cognitive ability, SES, and genetic ancestry are N = 1,369, N = 1,380, and N = 1,391, respectively.

4. Main Analyses

The key question is whether the bivariate associations between genetic ancestry and outcomes follow ancestry or racial/ethnic identification. This question was explored in two ways. First, the full sample was examined with SIRE being statistically controlled (4.1). Second, subgroup analyses were conducted (4.3). The analyses complement one another. The full sample analysis provides more power. Additionally, it allows for a robust exploration of the relation between parental SES, cognitive ability, and genetic ancestry (4.2). The subsample analyses allow focus to be directed to specific groups to see if the full sample associations replicate in subsamples.

4.1. Relationship between Genetic Ancestry, Cognitive Ability, and Parental SES

The main regression analysis strategy consisted of fitting models with SIRE (Model 1) and SIRE and genetic ancestry (Model 2). As there were two outcomes (cognitive ability and parental SES) and four methods of coding for SIRE, there were eight tables. The results from the standard and continuous SIRE models are shown in Tables 7–10. The results based on the common SIRE and the dummy SIRE methods were largely redundant with respect to those based on, respectively, the standard SIRE and the continuous SIRE methods, and so are presented in the Supplementary File 2 (Tables S1–S4). The unstandardized beta coefficients are shown with both White SIRE and European ancestry as the reference classes (and, thus, unstandardized betas of 0). Since cognitive ability and SES scores are already standardized, these unstandardized betas represent a change in a standard deviation of cognitive ability/SES over a change in percentage of a given ancestry. The adjusted coefficient of determination, denoted as r^2-adj., is reported to facilitate model comparison. As individuals were assigned ancestry percentages for each genetic ancestry group, the sample sizes are the same for each component. Across genetic ancestries, the standard errors varied substantially as a function of variance in admixture. Since one of the objectives was to determine the relative utility of including genetic ancestries as a predictor, ancestry components with high standard errors were not dropped (we did this, however, for some subgroup analyses below, see: Section 4.3 for discussion).

Table 7. Regression model results for cognitive ability using standard SIRE

	Model 1				Model 2			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI Lower**	**CI upper**
SIRE: White	0				0			
SIRE: African American	−0.97	0.09	−1.15	−0.80	0.31	0.18	−0.06	0.67
SIRE: American Indian	-0.09	0.47	-1.02	0.84	0.41	0.46	-0.5-	1.32
SIRE: Asian	-0.02	0.10	-0.21	0.18	-0.07	0.19	-0.44	0.30
SIRE: Hispanic	−0.62	0.07	−0.75	−0.48	−0.17	0.10	−0.36	0.03
SIRE: Multi ethnic	−0.24	0.08	−0.40	−0.08	0.04	0.11	−0.18	0.25
SIRE: Other	−0.33	0.22	−0.76	0.10	−0.03	0.23	−0.48	0.42
SIRE: Pacific Islander	−0.76	0.24	−1.23	−0.29	0.02	0.31	−0.58	0.63
African					−1.58	0.20	−1.98	−1.19
Amerindian					−1.15	0.31	−1.76	−0.55
East Asian					0.06	0.19	−0.31	0.42
Oceanian					−4.81	1.27	−7.31	−2.31
Central Asian					0.21	0.26	−0.29	0.71
English as First Language	0.04	0.09	−0.13	0.21	0.03	0.09	−0.14	0.20

Note: $N = 1{,}369$ for both models. Model 1: r^2-adj. = 0.11; Model 2: r^2-adj. = 0.16.

Table 8. Regression model results for parental SES using standard SIRE

	Model 1				Model 2			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI Lower**	**CI upper**
SIRE: White	0				0			
SIRE: African American	-0.98	0.09	1.14	0.81	0.28	0.18	0.06	0.63
SIRE: American Indian	-1.45	0.46	2.34	0.55	0.85	0.44	1.71	0.02
SIRE: Asian	-0.17	0.09	0.36	0.01	0.03	0.18	0.38	0.32
SIRE: Hispanic	-0.84	0.06	0.97	0.71	0.21	0.09	0.39	0.03
SIRE: Multi ethnic	-0.46	0.08	0.61	0.31	0.03	0.10	0.23	0.18
SIRE: Other	-0.52	0.24	0.99	0.05	0.39	0.24	0.85	0.08
SIRE: Pacific Islander	-1.62	0.23	2.08	1.17	0.22	0.29	0.79	0.35
African					-1.55	0.19	1.93	1.18
Amerindian					-1.91	0.29	2.48	1.34
East Asian					-0.24	0.18	0.59	0.11
Oceanian					-7.23	1.20	9.59	4.87
Central Asian					0.62	0.25	0.14	1.11
English as First Language	0.13	0.08	−0.03	0.29	0.11	0.08	−0.05	0.27

Note: N = 1,380 for both models. Model 1: r^2-adj. = 0.17; Model 2: r^2-adj. = 0.25.

Table 9. Regression model results for cognitive ability using continuous SIRE

	Model 1				**Model 2**			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI Lower**	**CI upper**
SIRE: White	0				0			
SIRE: Hispanic	-0.92	0.10	-1.11	-0.72	-0.46	0.15	-0.76	-0.16
SIRE: Pacific Islander	-1.00	0.18	-1.34	-0.65	-0.78	0.31	-1.39	-0.17
SIRE: Asian	0.04	0.09	-0.14	0.21	-0.04	0.22	-0.48	0.40
SIRE: African American	−1.02	0.08	−1.18	−0.85	0.29	0.24	−0.18	0.75
SIRE: American Indian	−0.62	0.26	−1.13	−0.10	−0.24	0.27	−0.76	0.28
SIRE: Other	−0.35	0.22	−0.77	0.08	−0.04	0.23	−0.49	0.42
African					−1.59	0.27	−2.13	−1.06
Amerindian					−0.75	0.33	−1.40	−0.11
East Asian					0.06	0.23	−0.40	0.52
Oceanian					−1.57	1.50	−4.50	1.37
Central Asian					0.18	0.28	−0.37	0.72
English as First Language	0.05	0.08	−0.11	0.21	0.04	0.08	−0.13	0.20

Note: N = 1,369 for both models. Model 1: r^2-adj. = 0.15; Model 2: r^2-adj. = 0.17.

Table 10. Regression model results for parental SES using continuous SIRE

	Model 1				Model 2			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI Lower**	**CI upper**
SIRE: White	0				0			
SIRE: Hispanic	-1.12	0.09	-1.31	-0.94	-0.28	0.14	-0.56	0.00
SIRE: Pacific Islander	-2.09	0.17	2.41	1.76	2.06	0.29	2.63	-1.50
SIRE: Asian	-0.16	0.08	0.32	0.01	0.56	0.21	0.97	0.15
SIRE: African American	-1.00	0.08	1.16	0.85	0.42	0.23	0.03	0.87
SIRE: American Indian	-1.14	0.25	1.62	0.66	0.63	0.25	1.11	0.15
SIRE: Other	-0.53	0.23	0.98	0.08	0.46	0.24	0.92	0.01
African					-1.76	0.26	2.27	1.24
Amerindian					-1.88	0.31	2.48	1.28
East Asian					0.33	0.22	0.10	0.75
Oceanian					-1.60	1.38	4.32	1.11
Central Asian					1.07	0.27	0.55	1.59
English as First Language	0.14	0.08	−0.01	0.30	0.12	0.08	−0.03	0.28

Note: $N = 1{,}380$ for both models. Model 1: r^2-adj. = 0.22; Model 2: r^2-adj. = .27.

In general, the results of the four models yielded consistent patterns (for computations, see Tables S1–S4 in Supplementary File 2). Across all four models (the two sets reported above and the two reported in the supplementary file), African ancestry is consistently associated with large negative betas (cognitive ability: mean −1.46, range −1.31 to −1.59; SES: mean −1.47, range −1.23 to −1.76). Amerindian ancestry is consistently associated with large negative betas (cognitive ability: mean −0.92, range −0.75 to −1.15; SES: mean −1.90, range −1.76 to −2.06). Oceanian ancestry is consistently associated with negative betas (cognitive ability: mean: −2.76, range −1.32 to −4.81; SES: mean −4.02, range −1.58 to −7.23), though the upper and lower confidence intervals overlapped with zero, as shown in the supplementary file, so nothing definitive can be inferred here. Central Asian ancestry is associated with positive betas (cognitive ability: mean 0.17, range 0.15 to 0.21; SES: mean 0.81, range 0.62 to 1.07). East Asian ancestry did not show any consistent patterns (cognitive ability: mean 0.03, range −0.01 to 0.06; SES: mean 0.02, range −0.24 to 0.33). English as a first language is associated with consistent but weakly positive betas in the genetic ancestry models (cognitive ability: mean 0.04, range 0.03 to 0.05; SES: mean 0.13, range 0.11 to 0.16). These latter findings might reflect small effects of language bias in the tests, as well as the effects of factors related to parental immigration status. Finally, consistent with other research, continuous SIRE coding was a better predictor of cognitive ability and SES than was standard SIRE coding. And, as expected given the covariance between continuous SIRE and genetic ancestry, the use of continuous SIRE reduced the incremental predictive ability of genetic ancestry.

As the SIRE effects had wide confidence intervals, interpretation warrants caution. That said, the African American SIRE betas are universally positive across codings in the genetic ancestry model (cognitive ability: mean 0.24, range 0.09 to 0.31; SES: mean 0.26, range 0.02 to 0.42), with the exception of the "Hispanic, African American" common combination group (cognitive ability: −0.38; SES: −0.40). For Hispanic SIRE groups, the situation was reversed in that these predictors were uniformly related to worse outcomes (cognitive ability: mean −0.31, range −0.46 to −0.17; SES: mean −0.20, range −0.28 to −0.11). Since these SIRE effects are independent of genetic ancestry, they may reflect the effects of social favoritism/discrimination, or of SIRE-related cultural factors. Alternatively, they could be the product of selective ethnic attrition (e.g., Duncan & Trejo, 2018) and genetic differences resulting from ethnic leakage (see, e.g., Harpending & Cochran, 2015).

The results were robust to controls for MRI scanner location, a proxy for geographic location, with each site entered as a dummy variable (results not shown).

4.2. Relationship between Genetic Ancestry, Cognitive Ability, and Parental SES

Parental SES is a potential confound in models where children's cognitive ability is an outcome. This is particularly the case given the moderate heritability of cognitive ability in childhood/early adolescence. In this case, it could be argued that associations between genetic ancestry and cognitive ability are consequent of those between parental genetic ancestry and parental SES. Figure 1 shows a theoretical path model of the variables. (We did not run a path model for this diagram as we lack trans-ethnically valid cognitive ability PGS).

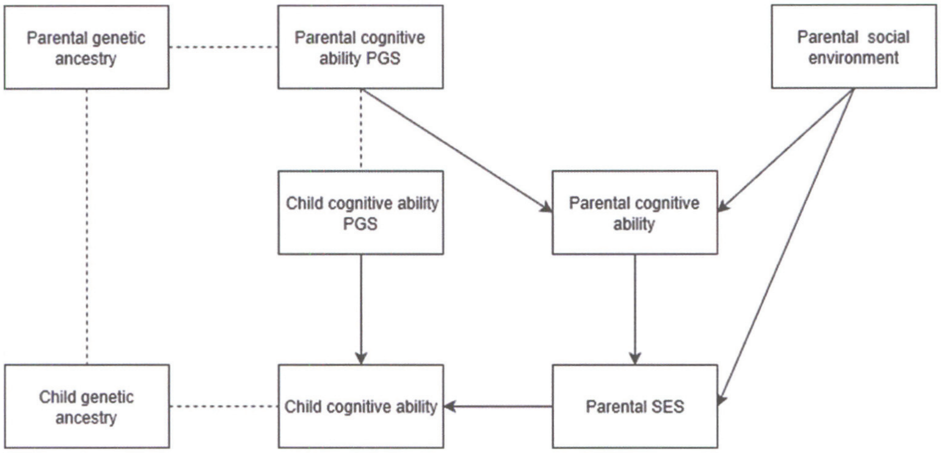

Parental genetic ancestry
Parental cognitive ability PGS
Parental social environment
Child cognitive ability PGS
Parental cognitive ability
Child genetic ancestry
Child cognitive ability
Parental SES

Figure 1. Theoretical path model for relationships between genomic ancestry, cognitive ability and parental SES

Accordingly, offspring ancestry is a function of parental ancestry and is correlated with offspring trait PGS. Likewise, (1) parental ancestry is correlated with parental trait PGS; (2) offspring PGS is a function of parental PGS; (3) parental trait is a function of parental PGS and social environment; (4) parental socioeconomic status is a function of both parental trait and social environment; and (5) offspring trait is a function of parental provided environment and parental provided genotype.

Without longitudinal data, disentangling causal pathways is impossible, since controlling for parental SES has the effect of controlling for an indeterminate portion of those behavioral traits directly transmitted by parents. It also controls for what is sometimes referred to as "genetic nurture" (Kong et al., 2018), which refers to genetic effects from parents on offspring that condition offspring traits by way of nurturing (e.g., the socioeconomic environment provided). This caveat noted, the regression results for standard SIRE are reported in Table 11 (Tables S5–S7 show the results for the common combination, dummy SIRE, and continuous method).

The betas for the genetic ancestries were reduced in size, especially in the case of Amerindian ancestry, but the directions were consistent with previous results.

For ease of summary, the relative importance of the predictors across ancestries and across the four models was quantified by calculating etas from ANOVA using sum of squares type II (following the methods detailed in Supplementary File 2, Appendix 1). Before the inclusion of parental SES, the mean total eta for the independent effect of genetic ancestry on cognitive ability was 0.18; after the inclusion of parental SES it was reduced to 0.11. However, it still outperformed SIRE in all models (means of 0.26, 0.11, 0.08, and 0.00 for parental SES, genetic ancestry, SIRE, and language, respectively). This is illustrated in Figure 2. (See Tables S19–S21 for complete results.)

Table 11. Regression model results for cognitive ability using standard SIRE and including SES

	Model 1				Model 2			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI Lower**	**CI upper**
(Parental) SES	0.35	0.03	0.29	0.40	0.31	0.03	0.26	0.36
SIRE: White	0				0			
SIRE: African American	−0.62	0.09	−0.80	−0.45	0.19	0.18	−0.16	0.54
SIRE: American Indian	0.42	0.45	−0.46	1.29	0.66	0.44	−0.21	1.53
SIRE: Asian	0.05	0.09	−0.13	0.23	−0.08	0.18	−0.43	0.28
SIRE: Hispanic	−0.32	0.07	−0.45	−0.19	−0.11	0.10	−0.30	0.08
SIRE: Multi ethnic	−0.08	0.08	−0.23	0.07	0.03	0.11	−0.17	0.24
SIRE: Other	−0.01	0.23	−0.47	0.45	0.14	0.24	−0.33	0.61
SIRE: Pacific Islander	−0.20	0.23	−0.64	0.25	0.09	0.29	−0.49	0.66
African					−1.05	0.20	−1.44	−0.66
Amerindian					−0.53	0.30	−1.12	0.05
East Asian					0.16	0.18	−0.20	0.51
Oceanian					−2.62	1.23	−5.04	−0.20
Central Asian					0.04	0.25	−0.45	0.53
English as First Language	0.00	0.08	−0.16	0.15	0.00	0.08	−0.16	0.15

Note: $N = 1,363$ for both models. Model 1: r^2-adj. = 0.21; Model 2: r^2-adj. = 0.23.

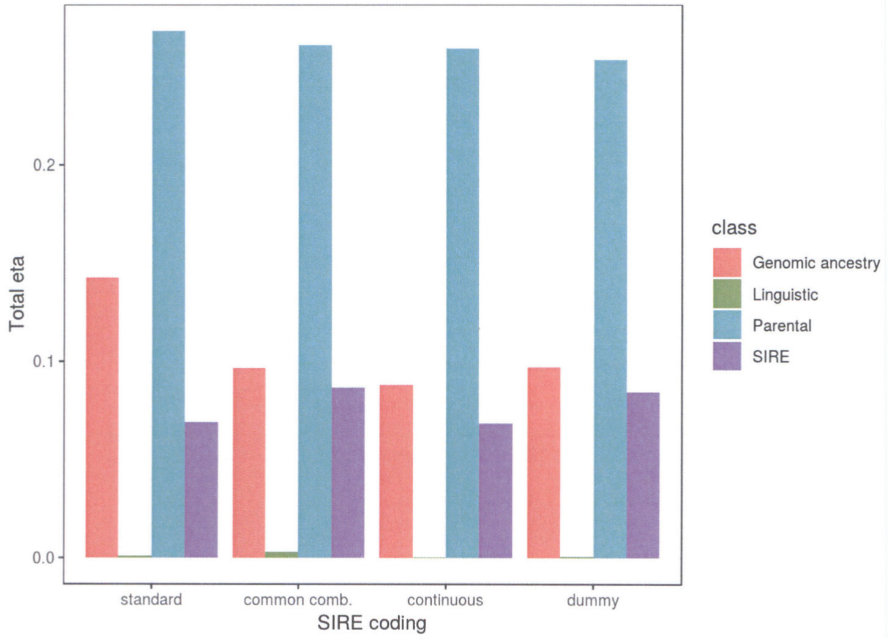


Figure 2. The median total eta for the independent effect of genetic ancestry on cognitive ability compared to SIRE for all models after controls for parental SES, genetic ancestry, SIRE, and language.

4.3. Subgroup Analyses

A number of subgroup analyses were carried out to see whether the results hold when data are disaggregated by SIRE groups. For this purpose, three subgroups of interest were chosen: (1) non-Whites, (2) African Americans, and (3) Hispanics. Models for other subgroups were not fit as the samples were too small to allow for an analysis with sufficient power to detect differences given the predictions of tenable evolutionary hypotheses or because, in the case of non-Hispanic mono-SIRE Whites, there was little variance in admixture.

4.3.1. Non-Whites

For this analysis, everyone classified as standard coding White was excluded. This group corresponds to the sociological construct of "people of color." Dummy coded SIRE predictors were used to avoid having to choose a new reference class, since doing so would make the betas less comparable with the regressions previously noted. A White SIRE predictor is included as some multiracial individuals identify as partially White. Table 12 shows the regression results. For Model 1 and Model 2 the dependents are cognitive ability and SES, respectively. The results are similar to those found previously, except that Oceanian ancestry was not a robust predictor here. Supplementary Table S8 shows the correlation matrix.

Table 12. Regression results for Non-Whites with cognitive ability (Model 1) and parental SES (Model 2) as dependents

	Model 1: Cognitive Ability				**Model 2: SES**			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI lower**	**CI upper**
SIRE: White	0.20	0.09	0.02	0.38	0.28	0.09	0.11	0.45
SIRE: Hispanic	−0.20	0.10	−0.40	0.00	−0.16	0.10	−0.35	0.03
SIRE: Pacific Islander	−0.45	0.14	−0.72	−0.18	−1.00	0.13	−1.25	−0.74
SIRE: Asian	0.20	0.13	−0.06	0.46	0.04	0.13	−0.21	0.29
SIRE: African American	0.30	0.16	−0.02	0.61	0.28	0.16	−0.03	0.59
SIRE: American Indian	−0.12	0.13	−0.36	0.13	−0.21	0.12	−0.45	0.02
SIRE: Other	0.16	0.25	−0.33	0.64	−0.17	0.26	−0.69	0.34
African	−1.47	0.26	−1.98	−0.96	−1.40	0.26	−1.91	−0.90
Amerindian	−0.81	0.34	−1.48	−0.15	−1.68	0.33	−2.32	−1.04
East Asian	0.00	0.21	−0.41	0.41	0.02	0.20	−0.38	0.42
Oceanian	−1.34	1.37	−4.02	1.34	−1.53	1.31	−4.10	1.04
Central Asian	0.10	0.27	−0.44	0.64	0.70	0.27	0.17	1.23
English as First Language	0.09	0.10	−0.10	0.29	0.28	0.09	0.10	0.46

Note: $N = 802$ for CA: r^2-adj. = 0.18; $N = 810$ for SES: r^2-adj. = 0.25.

4.3.2. African Americans

For African Americans, two inclusion criteria were utilized, yielding a narrow and a broad subsample. For the narrow (standard SIRE) subsample, only those who self-identified as African Americans and no other SIRE were included. For this analysis, only one ancestry predictor, African, was included as there was low variance for the other non-European genetic ancestries, and since this model showed a relatively good fit.

For the broad coding (in which everyone who identified as "African American" was included), it was an open question as to which model to use. The general criterion for model selection was adjusted r^2. In case of similar model fits, European ancestry was left out, the simplest models were used (in order to reduce standard errors of the predictors), and models which showed symmetry between the two outcomes were preferred (to ease comparison). Results for chosen models are shown in Tables 13 and 14 (the remainder can be found in the study notebook).

Table 13. Regression results for African Americans (Standard Coding) with cognitive ability (Model 1) and parental SES (Model 2) as dependents

	Model 1: Cognitive Ability				**Model 2: SES**			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI lower**	**CI upper**
African	−1.60	0.61	−2.81	−0.39	−2.50	0.65	−3.79	−1.22

Note: $N = 138$ for CA: r^2-adj. = 0.04; $N = 139$ for SES: r^2-adj. = 0.09.

Table 14. Regression results for African Americans (Broad Coding) with cognitive ability (Model 1) and parental SES (Model 2) as dependents

	Model 1: Cognitive Ability				**Model 2: SES**			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**CI lower**	**CI upper**
African	−1.24	0.27	−1.77	−0.71	−0.77	0.27	−1.30	−0.24

Note: $N = 225$ for CA: r^2-adj. = 0.08; $N = 227$ for SES: r^2-adj. = 0.03.

The betas for African ancestry were consistently negative. Tables S9 and S10 (Supplementary File 2) show the correlation matrices for genetic ancestry, SES, and cognitive ability. Figure 3 below shows the plot of European ancestry × cognitive ability in the standard African American population (in turquoise) along with the distribution of ancestry in the standard White population (in red). The large circles represent population means.

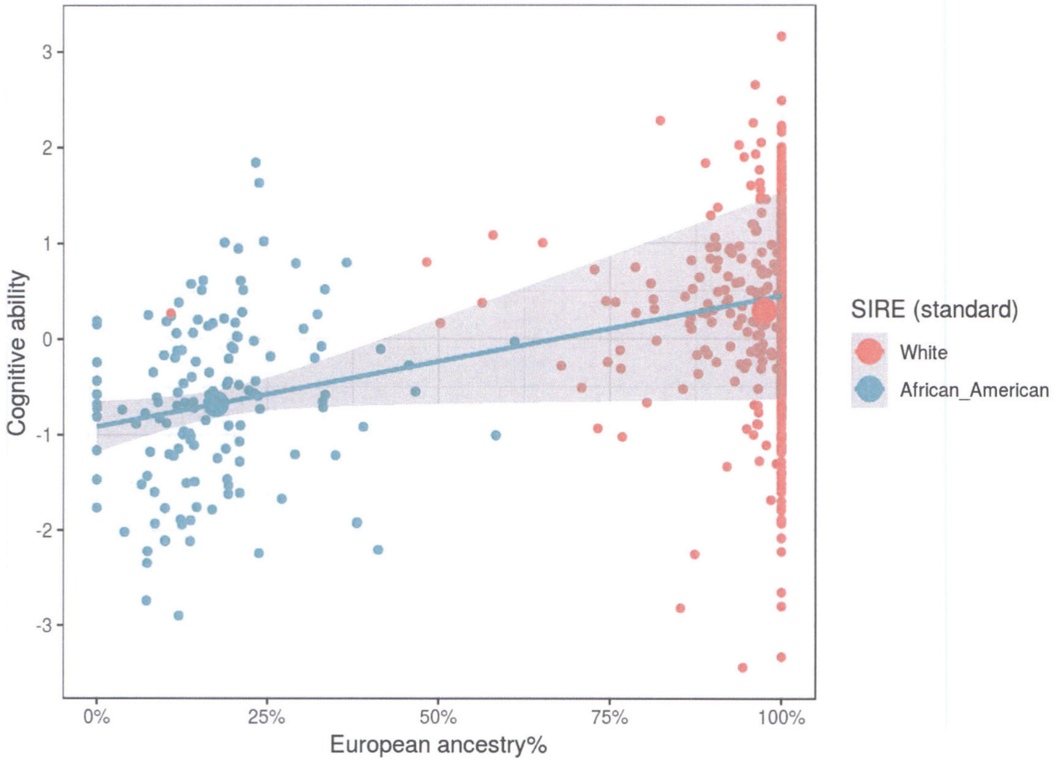


Figure 3. Association between European ancestry and cognitive ability (z score) in the African American (Standard) SIRE Group

4.3.3. Hispanics

For Hispanics, consistent with US research practices, every participant who self-identified as having Hispanic ethnicity was included. Four ancestry components showed sufficient variation to possibly warrant inclusion (African, Amerindian, East Asian, and European). European ancestry was dropped since it did not improve model fit (inclusion led to multicollinearity). The regression results are shown in Table 15.

Table 15. Regression results for Hispanics (standard coding) with cognitive ability (Model 1) and parental SES (Model 2) as dependents

	Model 1: Cognitive Ability				Model 2: SES			
Predictor	**Beta**	***SE***	**CI lower**	**CI upper**	**Beta**	***SE***	**Predictor**	**Beta**
African	-1.62	0.30	-2.22	-1.03	-2.00	0.26	-2.52	-1.48
Amerindian	-1.16	0.34	-1.84	-0.49	-2.09	0.30	-2.69	-1.49
East Asian	−0.53	0.28	−1.09	0.02	−1.32	0.25	−1.81	−0.82

Note: $N = 323$ for CA: r^2-adj. = 0.08; $N = 328$ for SES: r^2-adj. = 0.19.

The betas for both African and Amerindian ancestry were strong for this subsample. Interestingly, first language had little predictive ability, even though it would be expected to have more for the Hispanic sample, which likely contains some children of recent immigrants. Table S11 shows the correlation matrix. Figure 4 replicates Figure 3 but with Hispanics instead of African Americans.

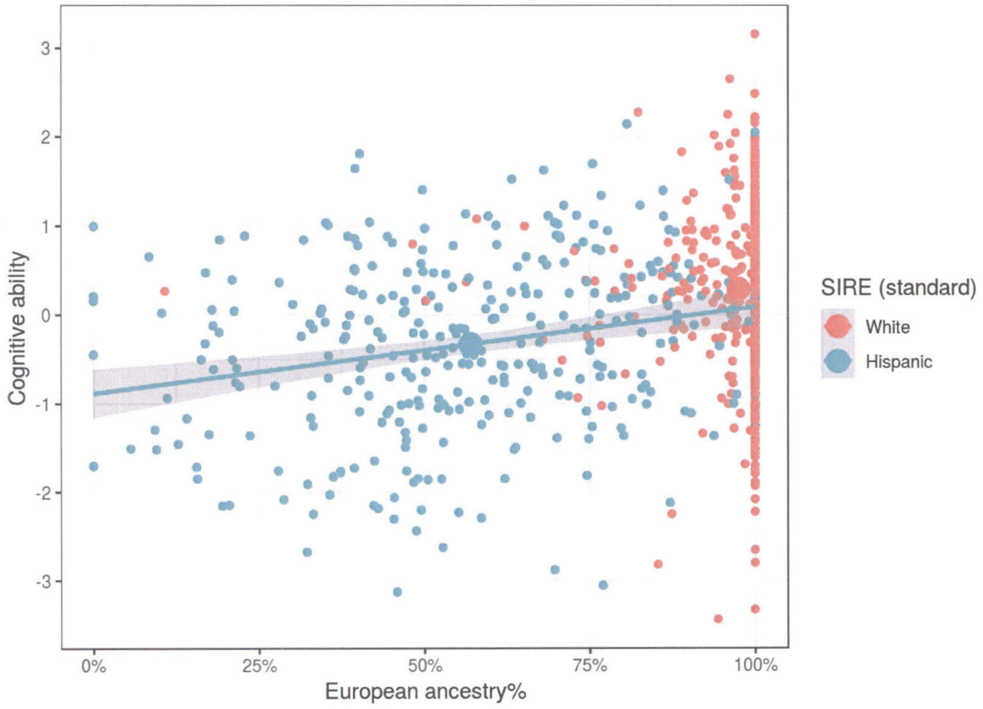


Figure 4. Association between European ancestry and cognitive ability (z score) in the Hispanic American (Standard) SIRE group

5. Discussion and Conclusions

Our results show that independent of SIRE, African, Oceanian, and Amerindian ancestries, relative to Eurasian ones, were associated with lower cognitive ability and parental SES. Relative to European ancestry, there were no clear associations for East Asian ancestry. East Asians in the US are a heterogeneous group, comprised of both South East Asians (e.g., Filipinos and Vietnamese) and North East Asians (e.g., Chinese, Japanese, and Koreans). Since at the national level these groups have different cognitive ability levels (Lynn, 2015), it is difficult to interpret these particular results in light of any global theory of cognitive differences.

Genetic ancestry was found to be related to both cognitive ability and parental socioeconomic status independent of SIRE, as predicted by evolutionary theory. These findings strongly disconfirm claims made by various researchers that there are no statistical relationships between genomic ancestry and cognitive ability when controlling for socially identified racial groups (Colman, 2016; Nisbett, 2009; Nisbett et al., 2012; Sternberg, 2005). Although these findings are congruent with predictions from evolutionary-genetic models, it should be kept in mind that genomic ancestry may also be associated with a number of non-genetic variables that run in families for environmental reasons, some of which may be causal. As such, the apparent validity of genetic ancestry could be due to confounding with these non-genetic variables, or could reflect ancestry-induced social processes. Still, it is worth noting that our results are especially striking in light of the stronger effect of environmental influences on younger compared to older persons' cognitive ability (Bouchard, 2018). Given this fact, it would be reasonable to expect that at least some environmental factors related to SIRE have their largest potential effects on cognitive ability among young individuals. Yet, in our sample of children, genetic ancestry explained a very great deal of inter-BGA cognitive ability variation, net of SIRE, potentially indicating that environmental effects in general have a limited role in intra-national cognitive ability differences between BGAs.

One set of the analyses included parental SES as a predictor of children's cognitive ability, and it was found to be a useful predictor. Inclusion of this in the models reduced the validity of ancestry predictors by 36–48% (cf. Section 4.2). As discussed above, a reduction in the effect size is expected on both genetic and non-genetic models. This is because in genetic models, controlling for parental SES indirectly controls for parental cognitive ability, parental genetic influences on cognitive ability, and ultimately children's genetic influences on cognitive ability, and thus the association between genomic ancestry and the relevant genetic factors is weakened.

Overall, genetic ancestry was a better predictor of outcomes than was SIRE membership. As the US population becomes more ancestrally heterogeneous (owing to admixture), and SIRE and genetic ancestry become less related, genetic ancestry may turn out to be an even better tool for studying race-related social differences. Research suggests this may be the case for very-admixed countries such as Brazil (e.g., Da Silva et al., 2018).

5.1. Non-Evolutionary Explanations

While the results we found are consistent with an evolutionary model, there are some potential alternative explanations, namely: phenotypic discrimination, confounding due to immigration status, confounding due to geographic location, and intergenerational environmental transmission.

Some have posited discrimination based on stereotypical race-phenotype, called colorism (Telles, 2014). It has been argued that such discrimination could account for covariances between BGA, cognitive ability, and SES (e.g., Conley & Fletcher, 2017). Unfortunately, this dataset does not have appearance data (e.g., skin color), so we could not test whether the associations found are statistically mediated by phenotypic differences. With regard to this sample specifically, we find it unlikely that colorism could directly lead to the association between ancestry and cognitive ability given the ages of the participants. This is because most colorism models propose market-based discrimination (e.g., Marira & Mitra, 2013). A theoretical possibility is that such discrimination induces associations between parental SES and BGA and that parental SES differences influence offspring cognitive ability. However, some of our associations were only partially reduced in strength when controlling for parental SES, downgrading the likelihood of this scenario.

More generally, it is not clear that colorism is actually a potent force, at least in the USA. Consider research based on sibling designs, which can distinguish between discriminatory and intergenerational effects. A number of studies in the economics literature have utilized sibling control designs in this fashion (Francis-Tan, 2016; Kizer, 2017; Marteleto & Dondero, 2016; Mill & Stein, 2016; Rangel, 2014; Telles, 2006). Unfortunately, they differ somewhat in design (e.g., raw vs. SES-controlled results for between-family regressions), and do not report standardized effect measures, so we were unable to quantitatively meta-analyze them. However, generally speaking, when family characteristics are controlled for, residual associations between racial appearance and social outcomes are small. In the words of one researcher who studied a large dataset from Brazil: "[T]he estimated coefficients are small in magnitude, implying that individual discrimination is not the primary determinant of interracial disparities. Instead, racial differences are largely explained by the family and community that one is born into" (Francis-Tan, 2016). Mill and Stein (2016) make statements to the same effect based on an analysis of a large dataset from the USA.

Another possibility is confounding due to immigration status. Some SIRE groups in our study (specifically East Asians and Hispanics) contain substantial numbers of new immigrants. For these, possible interactions between generational status and admixture complicate interpretations of associations between BGA and social outcomes. For example, the US Hispanic population is comprised of ongoing waves of migrants primarily from Central America and the Caribbean. Since there is an association between immigrants' generational status and social outcomes (see Rindermann & Thompson, 2014), if there is likewise an association between ancestry and generation status, this could lead to biases in parameter estimates. It is difficult to untangle these effects without detailed information about migrant status and specific population histories. That said, Kirkegaard et al. (2017) showed that associations between SES and ancestry can be found across the Americas. It seems unlikely that Amerindian ancestry would be related to SES among native Mexicans, and that African ancestry would be related to SES among native Puerto Ricans, but that in the USA the associations within Latin-American-origin populations would only be due to migrant status confounding. And it also seems unlikely that the cause of the association between ancestry and cognitive ability would be radically different from the cause of the association between ancestry and SES. Furthermore, migrant status confounding cannot non-negligibly factor into associations with African ancestry found in our sample, indicating that non-genetic environmental explanations that rely on migrant status confounding would not be parsimonious.

Regarding geography, we were unable to directly control for the specific locations of the participants. However, when we ran the analyses adding controls for MRI scanner location, as a proxy for geographic location, there was no substantive effect. Overall, the three explanations just discussed seem unlikely to account for our findings. However, the potential confounds that they invoke warrant investigation in future research.

5.2. Related Research

While few studies have looked at the association between genetic ancestry and cognitive outcomes, a large number of older ones have examined the relationship between genealogical and phenotypic indexes of ancestry and cognitive performance. Summarizing research on the relationship between indexes of Amerindian ancestry and outcomes, Berry (1968) notes:

> Nevertheless, many other researchers have explored the relationship between academic achievement and certain indices of assimilation and have reached the same conclusion. Coombs (123:6) reports: "Amazingly consistent relationship between the degree of Indian blood and the pre-school language on the one hand and level of achievement on the other. These two characteristics are the best indices of the degree of acculturation.
>
> Atkinson (22) tested students at Union High School, Roosevelt, Utah, and found whites superior, mixed-bloods second, and full-bloods third—a fact frequently encountered in the literature. As many writers have pointed out . . . such terms as full-blood and mixed-blood refer to social rather than biological groups.

Berry (1968) and many of the researchers he cites interpret the apparent outcome-by-admixture associations as owing to cultural factors; notably, they also interpret the purported ancestry divisions as being delineated culturally, not genetically. Their view suggests that the factors relevant to performance differences can readily diffuse horizontally across cultural groups. Our results do not exclude this possibility but suggest that the relevant factors are being vertically transmitted along genealogical lines – the most plausible candidate being genes.

Loehlin, Lindzey, and Spuhler (1975) summarized research on the relationship between phenotypic indexes of African ancestry and outcomes:

> The majority of the studies with persons of mixed African-European ancestry found that groups of subjects judged to be of more African ancestry were on the average slightly inferior on the tests of intellectual functioning employed.
>
> Because of the probability of complex and differential environmental response to physical differences and the likelihood of assortative mating complicating the genetic picture, in addition to the questionable reliability of the racial measures themselves, it is easy to decide that these findings lend themselves to no firm conclusions. One might go beyond this generalization and suggest that the observed findings provide little solace for extremes of either environmentalism or genetical determinism.

Our research design addresses many of the problems noted by Loehlin et al. (1975). First, we use a multilocus index of genetic ancestry, so cross-assortative mating for specific phenotypic characters (such as skin color) and cognitive ability is not a possible confound. Next, our index of ancestry is highly reliable, unlike the blood group studies that Loehlin et al. (1975) discuss. Moreover, we show in our supplementary simulation that, owing to range restriction in ancestry, the typically small correlations are consistent with large between-group differences. Since there is disagreement about the interpretation of these pre-genomic studies, a systematic meta-analysis is warranted to confirm that indexes of ancestry are generally correlated with cognitive ability.

A number of studies have addressed the issue of the association between genetic ancestry and cognitive ability in more indirect ways (e.g., Rushton & Jensen, 2005). A popular approach involves testing Spearman's hypothesis, which posits that the magnitude of measured cognitive ability gaps between BGAs is a function of the *g* loadings of the test instruments (studies generally support Spearman's hypothesis; see Frisby & Beaujean, 2015; te Nijenhuis & Van den Hoek, 2016; Warne, 2016). The extent to which *g* moderates the magnitude of inter-BGA cognitive ability gaps may be relevant insofar as the *g* loadings of tests potentially correlate perfectly with their heritabilities (Van Bloois et al., 2009). Nonetheless, some argue that the truth of Spearman's hypothesis does not exclude the possibility that inter-BGA cognitive ability gaps have non-genetic environmental origins (e.g., Flynn, 2000). Future admixture analyses could resolve this debate over the implications of the truth of Spearman's hypothesis. For example, if the present model were run for separate subtests, the resultant vector of model beta values could be correlated with the vector of subtest *g* loadings to determine whether more *g*-loaded subtests involve more strongly genetic inter-BGA ability gaps.

5.3. Evolutionary-Genetic Explanations versus Familial Environmental Models

In the context of admixture analyses of morphological and health-related traits, it is generally accepted (e.g., in genetic epidemiology) that the finding of a correlation between outcomes and genetic ancestry constitutes support for evolutionary-genetic models of inter-BGA differences, especially if potential environmental confounds are statistically controlled (see Introduction). Logically the same inference should apply to behavioral traits and also to outcomes under partial genetic influence, such as educational attainment. (Others make the same point: Nisbett (2009) suggests that a relation between ancestry and IQ within SIRE groups would constitute "direct evidence" supporting hereditarian models of inter-BGA cognitive differences.) But many researchers are inconsistent in their treatment of these outcomes (see, e.g., Da Silva et al., 2018), falling for the so-called sociologist's fallacy (Flynn, 2012; Sesardic, 2005). (Specifically, the sociologist's fallacy refers to the hasty inference—made without considering the possible role of genetic factors—that a correlation between a social variable (such as SES) and a phenotype (such as cognitive ability) implies that the social variable is causal with respect to the phenotype.)

Support for evolutionary-genetic models provided by global admixture analysis is indirect because there are possible confounds, as discussed above. It is also possible that non-genetic familial models (as noted) make the same predictions. Evolutionary-genetic models, though, can be directly tested using admixture mapping (González et al., 2005), which in genetic epidemiology is usually the next step taken after consistent findings of an association between outcomes and global genetic ancestry.

The idea behind the method is fairly simple: though we may not know exactly where the causal variants are on the genome, we know roughly where they are located (i.e., in which genes), and they can plausibly be assumed to be the same loci across populations (Zanetti & Weale, 2016). This means that if we take cases and controls (or high/low trait groups) from a given admixed population where one ancestry (A) has a higher polygenic score for a trait (e.g., higher disease risk), the cases will have an overall higher proportion of that A ancestry than the controls. However, this increase in ancestry will not be randomly distributed along the genome but will be concentrated in regions around the causal variants. Thus, the test for evolutionary-genetic models is whether such local ancestry enrichment can be found for variants identified via GWASs for cognitive ability. As cognitive ability is strongly polygenic, this test will likely require a large sample size. This method can also be used to test racial-phenotypic discrimination models, as these models predict that associations between genetic ancestry and both behavioral and social outcomes will be more pronounced on regions of the genome related to visible phenotypes. Such analyses have been conducted in relation to assortative mating (Zou et al., 2015).

Following methods in genetic epidemiology, we began with a preliminary global BGA analysis, and have shown that non-White (except East Asian) ancestry in the United States is correlated with lower cognitive ability and parental SES. Before drawing conclusions, however, it is important to replicate these results. A single dataset or study cannot settle a debate (Schmidt & Oh, 2016), owing to potentially unidentified confounds. However, if our results can be replicated, admixture mapping, which can discriminate between phenotypic discrimination, environmental familial, and evolutionary-genetic models, is the next warranted step.

5.4. Implications for Research on Differences

Insofar as there is interest in ameliorating ethnic and racial differences in important behavioral and social outcomes, it is important to understand their causation. Results in this paper are largely in agreement with a familial model of differences in the Americas. These results suggest that outcome differences are being passed on along family lines, and that they are not due to common ethno-cultural factors. This is consistent with an evolutionary-genetic model, but environmental familial models cannot be ruled out at this stage.

These results clearly indicate that using standard categorical SIRE classifications fails to capture the full extent of BGA-associated disadvantages. As for other general models of American race/ethnic differences, contra Goldsmith et al. (2006), these results suggest the possibility of non-trivial omitted variable bias in discrimination-based models. Proponents of such models should try to show that their associations are independent of genetic ancestry. It is not clear that this is generally the case. (For example, see the 1982 Pelotas-birth cohort results reported in Supplementary File 2 of (Kirkegaard et al., 2017), which show no consistent association between interviewer-reported color and social outcomes after genetic ancestry is statistically controlled.) Proponents of cultural identification models are also encouraged to take genetic ancestry into account.

Supplementary Materials: An online supplementary file for this paper is available through Open Science Framework: https://osf.io/tvb5n/.

Author Contributions: E.O.W.K. and J.F.: Conceptualization; Methodology; Formal Analysis; Writing – Original Draft Preparation; Writing – Review & Editing; Project Administration; Data Curation. G.M.: Obtained Data Access; Writing – Review & Editing. R.L.W. – Original Draft Preparation; Writing – Review and Editing.

Funding: This research received no external funding.

Acknowledgments: The Pediatric Imaging, Neurocognition, and Genetics (PING) database (http://pingstudy. ucsd.edu) provided the data used in this study. Data collection and sharing were funded by NIH (National Institutes of Health) (grant RC2DA029475). PING data were provided by the PING Coordinating Center at the Center for Human Development, University of California, San Diego. PING is funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The manuscript "Biogeographic ancestry, cognitive ability, and socioeconomic outcomes" was approved by the PING consortium in the spring of 2017. Since then, the consortium head informed us that the PING dataset is no longer under PING consortium supervision and that the PING consortium no longer has the authority to review papers; he further noted that it is no longer necessary to list the PING as a co-author (6 September 2018). As such they are not listed. We contacted NIMH (National Institute of Mental Health), who now supervises the data, and they notified us that since we did not download the data through NIMH we were not bound by the NIMH user agreement nor did we need NIMH approval (7 September 2018). We notified them that we were planning on submitting the paper as approved by the PING consortium (14 September 2018), a notification they acknowledged (2 October 2018).

Conflicts of Interest: The authors declare no conflict of interest.

Supplementary Materials: An online supplementary file for this paper is available through Open Science Framework: https://osf.io/tvb5n/.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

References

Dolan, C. V., & Molenaar, P. C. (1994). Testing specific hypotheses concerning latent group differences in multi-group covariance structure analysis with structured means. *Multivariate Behavioral Research*, *29*(3), 203-222.

Duncan, L., Shen, H., Gelaye, B., Ressler, K., Feldman, M., Peterson, R., & Domingue, B. (2018). Analysis of polygenic score usage and performance across diverse human populations. *BioRxiv*, 398396.

Akshoomoff, N., Newman, E., Thompson, W. K., McCabe, C., Bloss, C. S., Chang, L., ... & Jernigan, T. L. (2014). The NIH Toolbox Cognition Battery: results from a large normative developmental sample (PING). *Neuropsychology*, *28*(1), 1.

Becker, D., & Rindermann, H. (2016). The relationship between cross-national genetic distances and IQ-differences. *Personality and Individual Differences*, *98*, 300-310.

Becker, N. S., Verdu, P., Froment, A., Le Bomin, S., Pagezy, H., Bahuchet, S., & Heyer, E. (2011). Indirect evidence for the genetic determination of short stature in African Pygmies. *American journal of physical anthropology*, *145*(3), 390-401.

Berry, B. (1968). *The education of the American Indians: A survey of the literature*. Columbus: Ohio State University. Available online: https://files.eric.ed.gov/fulltext/ED026545.pdf (accessed on May 17 2018).

Bidulescu, A., Choudhry, S., Musani, S. K., Buxbaum, S. G., Liu, J., Rotimi, C. N., ... & Gibbons, G. H. (2014). Associations of adiponectin with individual European ancestry in African Americans: the Jackson Heart Study. *Frontiers in genetics*, *5*, 22.

Bouchard, T. J., Jr. (2018). Hereditary ability: *g* is driven by experience-producing drives. In R. J. Sternberg (Ed.). *The nature of human intelligence*. United Kingdom: Cambridge University Press.

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of african americans, latinos, and european Americans across the United States. *The American Journal of Human Genetics*, *96*(1), 37-53.

Christainsen, G. B. (2013). IQ and the wealth of nations: How much reverse causality?. *Intelligence, 41*(5), 688-698.

Clark, G. (2007). *A farewell to alms: a brief economic history of the world*. Princeton University Press.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility: Surnames and the History of Social Mobility*. Princeton, NJ: Princeton University Press.

Cochran, G., & Harpending, H. (2009). *The 10,000 year explosion: How civilization accelerated human evolution*. New York: Basic.

Colman, A. M. (2016). Race differences in IQ: Hans Eysenck's contribution to the debate in the light of subsequent research. *Personality and Individual Differences*, *103*, 182-189.

Conley, D., & Fletcher, J. (2017). *The genome factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press.

Cottrell, J. M., Newman, D. A., & Roisman, G. I. (2015). Explaining the black–white gap in cognitive test scores: Toward a theory of adverse impact. *Journal of Applied Psychology, 100*(6), 1713–1736.

da Silva, T. M., Fiaccone, R. L., Kehdy, F. S., Tarazona-Santos, E., Rodrigues, L. C., Costa, G. N., ... & Barreto, M. L. (2018). Biogeographical ancestry is associated with socioenvironmental conditions and infections in a Latin American urban population. *SSM-population health*, *4*, 301-306.

Duncan, B., & Trejo, S. J. (2018). Identifying the later-generation descendants of US immigrants: Issues arising from selective ethnic attrition. *The ANNALS of the American Academy of Political and Social Science*, *677*(1), 131-138.

Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3801-3808.

Fan, C. C., Bartsch, H., Schork, A. J., Chen, C. H., Wang, Y., Lo, M. T., ... & Dale, A. M. (2015). Modeling the 3D geometry of the cortical surface with genetic ancestry. *Current biology*, *25*(15), 1988-1992.

Fedderke, J. W., Klitgaard, R. E., MacMurray, J., & Napolioni, V. (2014). Diagnosing deep roots of development: genetic, disease and environmental factors. *Economic Research Southern Africa*.

Flynn, J. R. (2000). IQ gains, WISC subtests and fluid *g*: *g* Theory and the relevance of Spearman's hypothesis to race. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The Nature Of Intelligence: Novartis Foundation Symposium, vol. 233* (pp. 202 – 227). Chichester, UK7 Wiley.

Flynn, J. R. (2012). *How to improve your mind: Twenty keys to unlock the modern world*. London: Wiley-Blackwell.

Francis-Tan, A. (2016). Light and shadows: an analysis of racial differences between siblings in Brazil. *Social Science Research*, *58*, 254-265.

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence*, *51*, 79-97.

Fuerst J. (2017). Lineage population: A concept needed by an observer of nature. M*ankind Quarterly,* *57*(4):590–631.

Fuerst, J., & Kirkegaard, E. O. (2016a). Admixture in the Americas: Regional and national differences. *Mankind Quarterly, 56*(3), 255.

Fuerst, J., & Kirkegaard, E. O. (2016b). The genealogy of differences in the Americas. *Mankind Quarterly, 56*(3), 425.

Giri, A., Edwards, T. L., Hartmann, K. E., Torstenson, E. S., Wellons, M., Schreiner, P. J., & Velez Edwards, D. R. (2017). African genetic ancestry interacts with body mass index to modify risk for uterine fibroids. *PLoS genetics*, *13*(7), e1006871.

Goldsmith, A. H., Hamilton, D., & Darity Jr, W. (2006). Shades of discrimination: Skin tone and wages. *American Economic Review*, *96*(2), 242-245.

González Burchard, E., Borrell, L. N., Choudhry, S., Naqvi, M., Tsai, H. J., Rodriguez-Santana, J. R., ... & Risch, N. (2005). Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. *American journal of public health*, *95*(12), 2161-2168.

Grall, L. (2016). *Ice age climate, somatic capital, and the timing of the neolithic transition* (No. 44-2016). MAGKS Joint Discussion Paper Series in Economics.

Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., ... & Yang, J. (2018). Global genetic differentiation of complex traits shaped by natural selection in humans. *Nature communications*, *9*(1), 1865.

Halder, I., Matthews, K. A., Buysse, D. J., Strollo, P. J., Causer, V., Reis, S. E., & Hall, M. H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep*, *38*(8), 1185-1193.

Harpending, H., & Cochran, G. (2015). Assortative Mating, Class, and Caste. In T. Shackelford & R. Hansen (Eds.), *The Evolution of Sexuality* (pp. 57- 67). New York: Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.

Kan, K-J. (2011). *The nature of nurture: The role of gene–environment interplay in the development of intelligence* (Ph.D. Dissertation). University of Amsterdam.

Kanazawa, S., 2012. *The Intelligence Paradox: Why the Intelligent Choice isn't Always the Smart One*. John Wiley & Sons, Inc., Hoboken, NJ.

Kirkegaard, E. O., & Fuerst, J. (2017). Admixture in Argentina. *Mankind Quarterly*, *57*(4), 542–580.

Kirkegaard, E.O.W.; Wang, M.; Fuerst, J. (2017). Biogeographic Ancestry and Socioeconomic Outcomes in the Americas: A Meta-Analysis. *Mankind Quarterly, 57*(3), 398–427.

Kizer, J. (2017). *Skin Color Stratification and the Family: Sibling Differences and the Consequences for Inequality*. UC Irvine. Available online: https://escholarship.org/uc/item/29n4r264 (accessed on July 28 2018).

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428.

Liebler, C. A., & Halpern-Manners, A. (2008). A practical approach to using Multiple-Race response data: A bridging method for publicuse microdata. *Demography*, *45*, 143-155.

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence*. New York: Freeman.

Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57-64.

Lynn, R. (2015). *Race differences in intelligence: An evolutionary analysis* (2nd ed.). Arlington, VA: Washington Summit Publishers.

Marigorta, U. M., & Navarro, A. (2013). High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS genetics*, *9*(6), e1003566.

Marira, T. D., & Mitra, P. (2013). Colorism: Ubiquitous yet understudied. *Industrial and Organizational Psychology*, *6*(1), 103-107.

Marteleto, L. J., & Dondero, M. (2016). Racial inequality in education in Brazil: A twins fixed-effects approach. *Demography*, *53*(4), 1185-1205.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., ... & Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499-503.

Meigs, J. B., Grant, R. W., Piccolo, R., López, L., Florez, J. C., Porneala, B., ... & McKinlay, J. B. (2014). Association of African genetic ancestry with fasting glucose and HbA 1c levels in non-diabetic individuals: the Boston Area Community Health (BACH) Prediabetes Study. *Diabetologia*, *57*, 1850-1858.

Mill, R., & Stein, L. C. (2016). Race, skin color, and economic outcomes in early twentieth-century America. *Available at SSRN 2741797*. Available online:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2741797 (accessed on June 9 2018).

Murray, C. (2007). The magnitude and components of change in the black–white IQ difference from 1920 to 1991: A birth cohort analysis of the Woodcock–Johnson standardizations. *Intelligence*, *35*(4), 305-318.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. New York, NY: WW Norton & Company.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Group differences in IQ are best understood as environmental in origin. *American Psychologist, 67*, 503-504.

Ntzani, E. E., Liberopoulos, G., Manolio, T. A., & Ioannidis, J. P. (2012). Consistency of genome-wide associations across major ancestral groups. *Human genetics*, *131*, 1057-1071.

Piffer, D. (2013). Factor analysis of population allele frequencies as a simple, novel method of detecting signals of recent polygenic selection: The example of educational attainment and IQ. *Mankind Quarterly*, *54*(2), 168-200.

Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence*, *53*, 43-50.

Racimo, F., Berg, J. J., & Pickrell, J. K. (2018). Detecting polygenic adaptation in admixture graphs. *Genetics*, *208*(4), 1565-1584.

Rangel, M. A. (2015). Is parental love colorblind? Human capital accumulation within mixed families. *The Review of Black Political Economy*, *42*(1-2), 57-86.

Revelle, W. (2017). *Psych: Procedures for Personality and Psychological Research*. Evanston, Illinois: Northwestern University. USA https://CRAN.R-p roject.org/package=psych.

Rindermann, H., Becker, D., & Coyle, T. R. (2016). Survey of expert opinion on intelligence: Causes of international differences in cognitive ability tests. *Frontiers in psychology*, *7*, 399.

Rindermann, H., Coyle, T., & Becker, D. (2013). 2013 survey of expert opinion on intelligence. Presented at the 14th Annual Meeting of the International Society for Intelligence Research, Melbourne, Australia, December 2013.

Rindermann, H., & Thompson, J. (2014). The cognitive competences of immigrant and native students across the world: An analysis of gaps, possible causes and impact. *Journal of biosocial science*, *48*(1), 66-93.

Rindermann, H., Woodley, M. A., & Stratford, J. (2012). Haplogroups as evolutionary markers of cognitive ability. *Intelligence*, *40*(4), 362-375.

Robinson, M. R., Hemani, G., Medina-Gomez, C., Mezzavilla, M., Esko, T., Shakhbazov, K., ... & Visscher, P. M. (2015). Population genetic differentiation of height and body mass index across Europe. *Nature genetics*, *47*(11), 1357-1362.

Rodriguez-Arana, A. (2010, September). Intelligence and the Wealth of Nations: Genetics Matter but there is Still Much Room to Reduce Inequalities preliminary. In *DEGIT Conference Papers* (No. c015_038). DEGIT, Dynamics, Economic Growth, and International Trade.

Roth, T. C., & Pravosudov, V. V. (2009). Hippocampal volumes and neuron numbers increase along a gradient of environmental harshness: a large-scale comparison. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1656), 401-405.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, public policy, and law*, *11*(2), 235.

Sayol, F., Maspons, J., Lapiedra, O., Iwaniuk, A. N., Székely, T., & Sol, D. (2016). Environmental variation and the evolution of large brains in birds. *Nature Communications*, *7*(1), 13971.

Schmidt, F. L., & Oh, I. S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology*, *4*(1), 32.

Sesardic, N. (2005). *Making sense of heritability*. Cambridge University Press.

Shriner, D. (2013). Overview of admixture mapping. *Curr. Protoc. Hum. Genet. Chapter 1*, 23.

Shriver, M. D., & Kittles, R. A. (2004). Genetic ancestry and the search for personalized genetic histories. *Nature Reviews Genetics, 5*(8), 611-618.

Sirin, S. R. (2005). The relationship between socioeconomic status and school outcomes [microform]: meta analytic review of research. *Rev Educ Res*, *75*(3), 417-53.

Snyderman, M., & Rothman, S. (1988). *The IQ controversy, the media and public policy*. Transaction Publishers.

Sol, D., Garcia, N., Iwaniuk, A., Davis, K., Meade, A., Boyle, W. A., & Szekely, T. (2010). Evolutionary divergence in brain size between migratory and resident birds. *PLoS One*, *5*(3), e9617.

Sol, D., Lefebvre, L., & Rodríguez-Teijeiro, J. D. (2005). Brain size, innovative propensity and migratory behaviour in temperate Palaearctic birds. *Proceedings of the Royal Society B: Biological Sciences, 272*(1571), 1433-1441.

Sol, D., Székely, T., Liker, A., & Lefebvre, L. (2007). Big-brained birds survive better in nature. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1611), 763-769.

Sternberg, R. J. (2005). There are no public-policy implications: A reply to Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11*, 295– 301.

te Nijenhuis, J., & Van den Hoek, M. (2016). Spearman's hypothesis tested on Black adults: A meta-analysis. *Journal of Intelligence*, *4*(2), 6.

Telles, Edward E. (2006). *Race in Another America: The Significance of Skin Color in Brazil*. Princeton, NJ: Princeton University Press.

Telles, Edward E. (2014). *Pigmentocracies: Ethnicity, Race, and Color in Latin America*. Chapel Hill, NC: UNC Press Books.

Templ, M., Kowarik, A., Alfons, A., & Prantner, B. (2019). VIM: Visualization and Imputation of Missing Values (Version 4.8.0). Available online: https://CRAN.R-project.org/package=VIM (accessed on 19 February 2015).

Templ, M., Kowarik, A., & Filzmoser, P. (2011). Iterative stepwise regression imputation using standard and robust methods. *Computational Statistics & Data Analysis*, *55*(10), 2793-2806.

Templeton, A. R. (2001). The genetic and evolutionary significance of human races. In J. Fish (Ed.), *Race and Intelligence: Separating science from myth* (pp. 31-56). Mahwah, NJ: Lawrence Erlbaum Associates.

Turchin, M. C., Chiang, C. W., Palmer, C. D., Sankararaman, S., Reich, D., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, & Hirschhorn, J. N. (2012). Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature genetics*, *44*(9), 1015-1019.

van Bloois, R. M., Geutjes, L. -L., te Nijenhuis, J., & de Pater, I. E. (2009, December 19). *g* loadings and their true score correlations with heritability coefficients, giftedness, and mental retardation: Three psychometric meta-analyses. *Paper presented at the Symposium on Group Differences, 10th Annual Meeting of the International Society for Intelligence Research, Madrid, Spain*.

van der Linden, D., Dunkel, C. S., & Madison, G. (2017). Sex differences in brain size and general intelligence (g). *Intelligence*, *63*, 78-88.

Wallace, A. R. (1864). The origin of human races and the antiquity of man deduced from the theory of "natural selection". *Journal of the Anthropological Society of London*, *2*, clviii-clxxxvii.

Warne, R. T. (2016). Testing Spearman's hypothesis with advanced placement examination data. *Intelligence*, *57*, 87-95.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Slotkin, J., ... & Gershon, R. (2014). The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: validation in an adult sample. *Journal of the International Neuropsychological Society*, *20*(6), 567-578.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in psychology*, 1832.

Winkler, C. A., Nelson, G. W., & Smith, M. W. (2010). Admixture mapping comes of age. *Annual review of genomics and human genetics, 11*, 65-89.

Woodley, M. A. (2010). Is Homo sapiens polytypic? Human taxonomic diversity and its implications. *Medical Hypotheses, 74*(1), 195-201.

Woodley, M.A., Figueredo, A.J., Sarraf, M.A., Hertler, S., Fernandes, H.B.F., & Peñaherrera-Aguirre, M. (2017). The rhythm of the west: A biohistory of the modern era, AD 1600 to the present. *Journal of Social Political and Economic Studies, monograph series, no. 37*. Washington DC: Scott Townsend Press.

Woodley, M. A., Younuskunju, S., Balan, B., & Piffer, D. (2017). Holocene selection for variants associated with general cognitive ability: Comparing ancient and modern genomes. *Twin research and human genetics*, *20*(4), 271-280.

Zanetti, D., & Weale, M. E. (2016). True causal effect size heterogeneity is not required to explain trans-ethnic differences in GWAS signals. *bioRxiv*, 085092.

Zou, J. Y., Park, D. S., Burchard, E. G., Torgerson, D. G., Pino-Yanes, M., Song, Y. S., ... & Zaitlen, N. (2015). Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. *Proceedings of the National Academy of Sciences*, *112*(44), 13621-13626

CHAPTER 6

GLOBAL ANCESTRY AND COGNITIVE ABILITY[1]

JORDAN LASKER, BRYAN J. PESTA, JOHN G.R. FUERST
AND EMIL O.W. KIRKEGAARD

Abstract

Using data from the Philadelphia Neurodevelopmental Cohort, we examined whether European ancestry predicted cognitive ability over and above both parental socioeconomic status (SES) and measures of eye, hair, and skin color. First, using multi-group confirmatory factor analysis, we verified that strict factorial invariance held between self-identified African and European-Americans. The differences between these groups, which were equivalent to 14.72 IQ points, were primarily (75.59%) due to difference in general cognitive ability (*g*), consistent with Spearman's hypothesis. We found a relationship between European admixture and *g*. This relationship existed in samples of (a) self-identified monoracial African-Americans (B = 0.78, $N = 2{,}179$), (b) monoracial African and biracial African-European-Americans, with controls added for self-identified biracial status (B = 0.85, $N = 2407$), and (c) combined European, African-European, and African-American participants, with controls for self-identified race/ethnicity (B = 0.75, $N = 7{,}273$). Controlling for parental SES modestly attenuated these relationships whereas controlling for measures of skin, hair, and eye color did not. Next, we validated four sets of polygenic scores for educational attainment (eduPGS). MTAG, the multi-trait analysis of genome-wide association study (GWAS) eduPGS (based on 8442 overlapping variants) predicted *g* in both the monoracial African-American ($r = 0.111$, $N = 2179$, $p < 0.001$), and the European-American ($r = 0.227$, $N = 4914$, $p < 0.001$) subsamples. We also found large race differences for the means of eduPGS ($d = 1.89$). Using the ancestry-adjusted association between MTAG eduPGS and *g* from the monoracial African-American sample as an estimate of the transracially unbiased validity of eduPGS (B = 0.124), the results suggest that as much as 20%–25% of the race difference in *g* can be naïvely explained by known cognitive ability-related variants. Moreover, path analysis showed that the eduPGS substantially mediated associations between cognitive ability and European ancestry in the African-American sample. Subtest differences, together with the effects of both ancestry and eduPGS, had near-identity with subtest *g*-loadings. This finding confirmed a Jensen effect acting on ancestry-related differences. Finally, we confirmed measurement invariance along the full range of European ancestry in the combined sample using local structural equation modeling. Results converge on genetics as a potential partial explanation for group mean differences in intelligence.

Keywords: race; intelligence; ancestry; socioeconomic status; discrimination; nature/nurture

---

[1] Editors' note: We have reformatted the original paper for this collection and updated the references to APA style.

1. Introduction

The European-American (EA) and African-American (AA) populations have been found to differ in mean general cognitive ability (or general intelligence, *g*) by about one standard deviation (Jensen, 1998; Roth et al., 2001; Shuey, 1966). This difference exists even though the heritability of intelligence is both high and virtually identical in the European and African-American populations (Pesta et al., 2019). Likewise, the shared and unshared environmental components of variance in *g* are similar in these groups. These findings, in conjunction with others (see, e.g., Jensen, 1998; Sesardic, 2005), have led some to propose the existence of a non-trivial genetic influence on group differences in cognitive ability (e.g., Jensen, 1969a, 1973, 1998; Lynn, 2015; Rushton & Jensen, 2005).

Researchers have devised various methods to assess the causality behind these differences (Jensen, 1969b; Rowe, 2005). Examples include transracial adoption studies (Scarr, 1977; Weinberg et al., 1992) (for a review, see: Thomas, 2017; Turkheimer, 1991), analysis of differential regression to the mean (Hu et al., 2019; Jensen, 1973; Murray, 1999), investigations of Spearman's hypothesis (Jensen, 1985; McDaniel & Kepes, 2014; te Nijenhuis et al., 2019), biometric decomposition of mean group differences (Jensen, 1998; Rowe & Cleveland, 1996), studies of multiracial persons (Loehlin et al., 1975; Rowe, 2005), and various intra- and cross-national comparisons (Lynn, 2015).

Some researchers have proposed the use of admixture analysis to assess the relationship between genetic ancestry and cognitive ability (Jensen, 1969b; Kirkegaard et al., 2019; Shockley, 1966; Warne, 2019). The premise of this approach is that among admixed population's individuals possessing different proportions of genetic ancestry from different populations will be more or less similar to members of these populations if they derive a greater or lesser proportion of their genetic ancestry from them. This simple genetic model predicts that individual resemblance to a given group mean in various outcomes (e.g., height, appearance, intelligence) will be a function of that individual's ancestry derived from that group (Kirkegaard et al., 2019; Templeton, 2001). When a difference in the level of some trait is observed between groups, an admixture analysis can provide indirect evidence that genetics plays a role in the difference.

A number of older studies have attempted to assess the relationship between ancestry and cognitive ability in admixed populations. Most of these studies, however, used indirect measures of ancestry, such as self-reports, genealogical records, or indices based on physical features like hair texture or skin color. These measures do not constitute objective or reliable indices of ancestry. As such, they show substantial measurement error and may thus be prone to giving inaccurate results. This may be especially true when the measures are based on phenotypic indicators of ancestry, as these are confounded with potentially biasing variables like discrimination, or cross-assortative mating for specific traits (Jensen, 1998; Reed, 1997; Shuey, 1966).

The above problem is common to all classical admixture studies. In addition to uncertainty regarding the extent of admixture, these studies also tend to suffer from small sample sizes, lack of replication, distinctly unrepresentative sampling, and young age at assessment (Rushton & Jensen, 2010). The issue of age at assessment can be severe, as one of the most well-replicated findings about the behavior genetics of intelligence is that heritability tends to increase with age (this has been dubbed the Wilson effect; Bouchard, 2013). As heritability rises, the shared environmental component of traits like intelligence tends to become insignificant (Bouchard & McGue, 2003; Briley & Tucker-Drob, 2013). If individuals from different races are reared in similar environments, typically observed differences in cognitive ability may not emerge until adulthood due to these shared environments masking them in childhood.

Modern DNA-based methods of assessing admixture are available, and with them researchers can generate accurate point estimates of individual ancestry from many different groups. Genomic admixture analysis has already proven to be a useful technique for investigating the causes of group differences in a variety of traits. For example, Becker et al. (2011) found that among Central African Pygmies, non-Pygmy admixture was associated with greater height. After the initial association between admixture and a trait is found to be robust, other methods like local admixture mapping are available to determine the specific genes and mechanisms behind group differences. Several authors (see, e.g., Jarvis et al., 2012; Migliano et al., 2013; Pemberton et al., 2018; Perry et al., 2014) have applied these techniques to Pygmies, finding that the Pygmy/non-Pygmy stature gap can be traced in part to specific differences in immune and neuroendocrine functioning.

Admixture analysis has been applied to many other traits, including diabetes (Cheng et al., 2012; Florez et al., 2009; Williams et al., 2000), hypertension (Kosoy et al., 2012), adaptation to high altitudes (Jeong et al., 2014), asthma (Flores et al., 2012), sleep depth (Halder et al., 2015), risk of prostate cancer (Freedman et al., 2006; Gudmundsson et al., 2007; Han et al., 2016; Yeager et al., 2007), and uterine fibroids (Giri et al., 2017). These studies generally find that admixture relates to the trait under investigation above and beyond covariates. Sometimes, however, the inclusion of putative environmental covariates, like diet or socioeconomic status (SES), removes the association with ancestry (e.g., Paiva, 2017). This suggests either that environmental covariates mediate the associations between ancestry and outcomes or that ancestry is a non-causal correlate of the trait in question that's actually caused by something else. This suggests either that environmental covariates mediate the associations between ancestry and outcomes or that ancestry is a non-causal correlate of the trait in question that's actually caused by something else if the mediation is not spurious. Ultimately, controls alone cannot establish causality (Pearl, 2012).

Four studies we know of have related genetically-assessed ancestry to cognitive ability in samples of African-Americans (Kirkegaard et al., 2019; Loehlin et al., 1973; Scarr et al., 1997; Warne, 2019). Others have done this for Hispanic Americans (Kirkegaard et al., 2019; Warne, 2019) and various groups in Latin America (Lima-Costa et al., 2016; Rothhammer & Llop, 1976; Valenzuela et al., 1998). Of these, the earlier ones (Loehlin et al., 1973; Rothhammer & Llop, 1976; Scarr et al., 1997; Valenzuela et al., 1998) relied on ancestry estimates derived from blood groups. The use of blood groups has been criticized on the grounds that they are unable to reliably assess global ancestry (Reed, 1997; Rowe, 2005).

A recent study by Kirkegaard et al. (2019) focused on the relationship between genetic ancestry and *g*. The authors analyzed the multisite Pediatric Imaging, Neurocognition, and Genetics (PING) dataset. They found that European ancestry was associated with *g* among both self-identified African and Hispanic Americans (respectively, $r = 0.20$, $N = 225$; $r = 0.23$, $N = 328$; Tables S10–S11). They assessed genetic ancestry with 15,000 SNPs, making their index highly-reliable (Ruiz-Linares et al., 2014). Ancestry explained both the full gap between African and European-Americans and most of the gap between Hispanic and European-Americans in *g*. However, the analysis had a few shortcomings. First, it was based on relatively small samples. Second, the heritability of *g* was unknown. Since the sample comprised children (with a mean age of 12 years) this is potentially concerning due to the Wilson effect. Third, strict factorial invariance (SFI) was not established so it cannot be unequivocally said that the group differences in *g* were measurement invariant. (In our supplementary materials we show that SFI along with the strong version of Spearman's hypothesis hold in the PING data.) Fourth, the sample was geographically dispersed. Such geographic dispersion could lead to spurious correlations between ancestry and cognitive ability (Conley & Fletcher, 2017), though the location of the MRIs did not affect estimates. Finally, as the authors note, no measure of skin color was included in the study, opening up the results to confounding by phenotype-based discrimination.

These results have been replicated in a preregistered study relating European ancestry to cognitive ability in a sample of African and Hispanic Americans ($r = 0.297$, $N = 193$; Warne, 2019). This sample was composed of adults, and used 27 items from the International Cognitive Ability Resource Team (Condon & Revelle, 2014) as a cognitive ability measure. This analysis also had shortcomings. For one, it was on a relatively small sample of individuals recruited online through Qualtrics (an online pollster) and participants self-reported their ancestry (as assessed by commercial DNA testing services such as Ancestry.com or 23andMe), opening these up to reporting error and bias due to differences in admixture estimates from different companies. Additionally, SES was not measured. Moreover, as with Kirkegaard et al. (2019), the heritability in this sample is unknown, there was no measure of skin color, and the sample was geographically dispersed.

The purpose of the present study is to conduct an admixture analysis capable of addressing the aforementioned problems with previous studies. We focus on African and European-Americans because the former represent a population substantially admixed with the latter. Generally, African-Americans have around 20% European DNA (Adhikari et al., 2017; Bryc et al., 2015; Cheng et al., 2012; Halder et al., 2008; McQueen et al., 2015). Provided that our African-American sample is sufficiently variable in terms of European admixture, we should be able to detect whether an association exists between cognitive ability and admixture from the two parent populations (i.e., Africans and Europeans) and, if so, to what degree genetic admixture mediates the observed mean difference between the groups. Moreover, if the samples are representative of their respective populations (and statistical power is sufficient), a finding of no association would indicate that an evolutionary-genetic explanation of the gap is unlikely (Nisbett, 2009; Templeton, 2001).

The present work uses a population-representative Philadelphia-based sample, the Philadelphia Neurodevelopmental Cohort (PNC), otherwise known as the Trajectories of Complex Phenotypes study (TCP). Due to the location, the results are directly comparable to those of Scarr et al. (1997). Our analysis has numerous advantages compared to earlier admixture studies. First, participants all came from the same location, so geographic confounding is not an issue. Second, we assessed measurement invariance (MI) for the cognitive test battery using multi-group confirmatory factor analysis (MGCFA; Dolan, 2000). Third, the heritabilities of the *g* factor and subtest scores have already been estimated for this sample. Specifically, Mollon et al. (2018) reported heritabilities for *g* of 0.61 (standard error (*S.E.*) = 0.14) and 0.72 (*S.E.* = 0.07) for the non-Hispanic African and European-Americans in this sample respectively. Fourth, we included estimates of skin, hair, and eye color to evaluate phenotypic discrimination (i.e., colorism) models of the observed differences. Fifth, we validated polygenic scores (PGS) associated with cognitive ability for both the African- and European-American samples and we examined to what extent cognitive ability- and education-related PGS (eduPGS), could account for group differences. Sixth, we tested for Jensen effects in relation to ancestry, heritability, and eduPGS. Seventh, we examined whether MI was tenable across the full range of European ancestry using local structural equation modeling (LSEM).

2. Materials and Methods

The TCP study was originally intended to evaluate behavioral dimensions predicting risk of mental illness (Calkins et al., 2015; Satterthwaite et al., 2016). The total sample includes data from 9421 genotyped participants assessed primarily from 2010 to 2013. Demographically, the sample was 51.7% female, 55.8% European-American, 32.9% African-American, and 11.4% Other, with a mean age of 14.2 (standard deviation (SD) = 3.7) years of age. Participants were recruited from the Philadelphia area. Persons with severe cognitive or medical impairments were excluded from the final sample. The subjects were English-speaking people aged 8–21 years at the time of testing. Access to the data was obtained through dbgap (phs000607.v3.p2) via Bryan J. Pesta.

2.1. Cognitive Ability

Participants were administered the Penn Computerized Neurocognitive Battery (PCNB; Moore et al., 2015). This battery was built to be highly-reliable, psychometrically-robust, and to incorporate tasks linked to specific brain systems. The PCNB consists of 14 tests grouped into five broad behavioral domains: Executive Control, Episodic Memory, Complex Cognition, Social Cognition, and Sensorimotor Speed. The tests in the battery are as follows: Penn Conditional Exclusion Test (PCET; meant to assess Mental Flexibility), Penn Continuous Performance Test (PCPT; Attention), Letter N-Back Task (LNB; Working Memory), Penn Word Memory Task (PWMT; Verbal Memory), Penn Face Memory Task (PFMT; Face Memory), Visual Object Learning Test (VOLT; Spatial Memory), Penn Verbal Reasoning Test (PVRT; Language Reasoning), Penn Matrix Reasoning Test (PMRT; Nonverbal Reasoning), Penn Line Orientation Test (PLOT; Spatial Ability), Penn Emotion Identification Test (PEIT; Emotion Identification), Penn Emotion Differentiation Test (PEDT; Emotion Differentiation), Motor Praxis Test (MP; Sensorimotor Speed), Finger Tapping (Tap; Sensorimotor Speed), and the Penn Age Differentiation Test (PADT; Age Differentiation). The sample also completed the Wide Range Achievement Test (WRAT), which is a highly-reliable broad ability measure (Satterthwaite et al., 2016).

Approximately 1.4% of the cognitive data were missing. We excluded individuals who did not have data for at least half the tests ($N = 115$). We imputed values for the remaining cases using IRMI (iterative robust model-based imputation) (Templ et al., 2011; 2019). This left us with complete cognitive data for 3,067 African- and 5,076 European-Americans. Since we are only concerned with African and European-Americans in this study we report results only for these groups. We residualized cognitive test variables for age and sex using a natural spline model.

In order to ensure this battery was unbiased, we followed the theoretical configuration reported by Moore et al. (2015) and performed an MGCFA (Dolan, 2000; Dolan & Hamaker, 2001; Dolan et al., 1994) with our two groups being self-identified African- and European-Americans. The group factors we used were Executive Functioning (EF), Episodic Memory (EM), and Complex Reasoning (CR). The tests which loaded onto these factors were the LNB, PCET, PCPT (EF), PFMT, PWMT, VOLT (EM), WRAT, PVRT, PLOT, and PMAT (CR). We attempted, alternatively a five factor model but the Social Cognition (PEIT, PEDT, PADT) and Sensorimotor Speed (MP, Tap) factors had low loadings and identification issues so we removed these tests.

The three-factor baseline model was found to fit well in both groups and the addition of a bifactor *g* considerably improved model fit. SFI was found for this model. The strong and contra forms of Spearman's hypothesis (Jensen, 1998) were rejected but the weak model fit the data well. Our constraint for testing the weak model was to set the episodic memory factor mean to 0 while the other group factor means were freed. The resulting mean differences appear in Table 1.

Table 1. Factor score differences between African and European-Americans based on the weak Spearman's Hypothesis model

Factor	Estimate	*SE*	Lower 95% CI	Upper 95% CI
g	1.046	0.020	1.007	1.085
Complex Reasoning	0.356	0.026	0.305	0.407
Executive Functioning Episodic Memory	−0.055	0.028	−0.109	−0.0001

Note: Positive values indicate higher European-American scores and vice-versa. SE = standard error.

These results accord with previous analyses (e.g., Frisby & Beaujean, 2015; Jensen, 1998) in that there is a nearly one *d* difference in cognitive ability between African and European-Americans and that this difference is concentrated on the *g* factor. The model without constraints for Spearman's hypothesis shows that (net of the *g* factor) there is an African-American advantage in executive functioning and episodic memory. This replicates patterns found in prior studies (e.g., Jensen, 1985; Frisby & Beaujean, 2015). The MGCFA results are discussed more fully in the Supplementary Materials. For the admixture analysis, we used *g* scores since differences were primarily relegated to this factor.

2.2. Parental Socioeconomic Status

Only paternal and maternal education were available to us as measures of SES. We coded parental education as the average of each, both *z*-scored first. When data were only available from one parent this value was used. The final score was then *z*-scored again. This variable was used as a control for SES in our admixture regressions but because we were unable to factor analyze a variety of SES measures and extract a more reliable indicator (Vyas & Kumaranayake, 2006) the control can only be regarded as partial. With this noted, parental education has been found to be more important than wealth in shaping children's intelligence (Rindermann & Ceci, 2018). The effect of SES as a background variable in our MGCFA is investigated in our Supplementary Materials.

2.3. Self-Identified Race/Ethnicity (SIRE)

Subjects chose their self-identified race/ethnicity (SIRE) based on yes/no questions wherein it was possible to select multiple races or ethnicities. The choices were: Black or African-American; American Indian or Alaskan Native; Asian; European-American; Hispanic/Latino; Native Hawaiian/Pacific Islander; Other; and Not Available/Pending Validation. We coded European-American following standard procedures: subjects who selected the European-American option only and did not report being Hispanic. African-American was coded the same way. Finally, we coded subjects who selected both European- and African-American as biracial African-European-American (AA|EA). Of the included participants, there were 5183 European-Americans, 3155 African-Americans, and 242 biracial African-European-Americans.

2.4. Genetic Ancestry Percentages

Satterthwaite et al. (2014) described the genotyping of the PNC dataset. Briefly, the participants were genotyped on six different arrays from two different manufacturers. The different arrays covered different variants. As such, it was necessary to impute them to obtain large, overlapping sets of single nucleotide polymorphisms (SNPs). We imputed the dataset using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html) with the Minimac3 imputation algorithm, 1000G Phase 3 v5 as reference panel, and Eagle v2.3 Phasing. In total, 15.5 M variants were available for analysis. To reduce computation time, we filtered these to variants with a minor allele frequency (MAF) of at least 0.05 in the subset of interest (i.e., the combined sample of European and African-Americans). That is, we selected only variants for which the second-most common alleles occurred in the population at a frequency of 0.05 or higher. This screening left 6. 5M variants to compute genetic ancestry percentages.

Note, this was only done in the context of computing admixture scores. We did not filter for MAF when computing color scores, as we did not want to lose variants that affected them. These operations were done using PLINK v1.90b6.8 (Chang et al., 2015). To estimate individual global ancestry we used ADMIXTURE version 1.3.0 D.H. (Alexander et al., 2009). Since we were only concerned with European and African-Americans, we ran ADMIXTURE with k = 2 genetic clusters. Some subjects either had no genotypes available or their data failed quality control. Thus, the final sample size was reduced to 7399. Consistent with previous research (e.g., Bryc et al., 2015), SIRE is strongly associated with genetic ancestry. This can be seen in Figure 1 which shows the probability of identifying with a particular SIRE group as a function of European admixture.

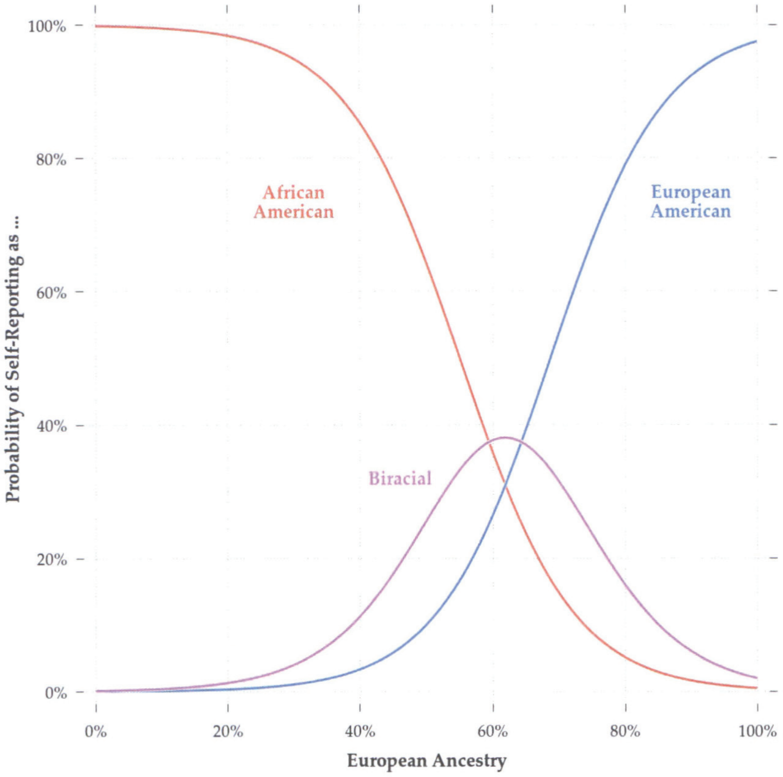


Figure 1. Logistic probability distribution for identifying as African-, biracial African European, and European-American as a function of genetic ancestry for the Trajectories of Complex Phenotypes (TCP) sample

2.5. Skin, Hair, and Eye Color

The data did not include measures of appearance, so we opted to impute these based on genotypes. We used the HIrisPlex-S web application to do this (https://hirisplex.erasmusmc.nl/). This application was developed by the U.S. Department of Justice for use in forensic investigations. It imputes skin, hair, and eye color probabilities with a high degree of accuracy based on 41 SNPs (with overlapping variants; 6 for eye color, 22 for hair color, and 36 for skin color). This tool has been validated on thousands of people from diverse regions of the world [79].

We focus on skin color since this trait is given primacy by colorist theorists (e.g., [80,81]) and because we were able to calculate skin color scores for a larger subsample than for hair or eye color. HIrisPlex-S gives probabilities of the Fitzpatrick Scale skin type (i.e., Type I, scores 0–6, "palest, freckles"; Type II, scores 7–13; Type III-IV (combined), scores 14–27; Type V, scores 28–34; Type VI, scores 35–36, "deeply pigmented dark brown to darkest brown"). We weighted the medium score of each Type (e.g., Type I = 3) by the probability of each type in order to create a single color measure. Unfortunately, because of the poor tagging of the relevant variants in some of the arrays, the imputed phenotypes were only available for a subset of the data ($N = 7321$ with cognitive ability and ancestry, versus $N = 5533$ with cognitive ability, ancestry, and skin color). By this scale, higher scores indicate darker colors.

The correlation between our imputed color score and European ancestry was −0.87 for the combined sample ($N = 5585$) and −0.39 for the African-Americans only sample ($N = 1557$). Figure 2 shows the regression plot for skin color and European ancestry, with SIRE groups marked in red (African-American), green (biracial African-European-American), and blue (European-American). The correlation found among African-Americans is similar to those previously reported (r's 0.25 to 0.71; median: ~0.50) for African-Americans and other groups of African descent (Halder et al., 2008; Leite et al., 2011; Parra et al., 2004; Ruiz-Linares et al., 2014). Furthermore, the predicted skin color values matched the expected values on the Fitzpatrick scale. The mean score for European-Americans was 14.70, placing them in category III (i.e., sometimes mild burn, tans uniformly) bordering on category II (i.e., usually burns, tans minimally), and for African-Americans, 30.96, placing them in category V (very rarely burns, tans very easily).

2.6. Cognitive Ability-Related Polygenic Scores

We scored the genomes using PLINK v1.90b6.8 (Chang et al., 2015) based on the genome-wide association study (GWAS) results for educational attainment reported in Lee et al. (2018). We constructed four different PGS based on the relative trade-off between differential linkage disequilibrium (LD) decay bias and the within-population signal. For background, a PGS "represents an aggregation, usually in the form of a sum, of the estimated effect sizes of the genetic variants in the genome" (Rosenberg et al., 2018). LD is the non-random association of alleles at different loci in a particular population. LD decay is the breakdown of this association between divergent populations. LD decay bias is bias in PGS—which are calculated based on both causal variants and variants in linkage with causal and non-causal ones—due to populations having different patterns of associations between variants and sites that are nearby (or in linkage).

First, we used the eduPGS with all variants trained without the 23andMe cohort ($N = 7{,}762{,}369$ SNPs overlapping in the present dataset) to construct a predictor with good within-population validity but very high LD decay bias (Duncan et al., 2018; Zanetti & Weale, 2018). Second, we used the multi-trait analysis of genome-wide association study (MTAG) eduPGS 10k SNPs ($N = 8442$ overlapping variants) to construct a predictor with good within-population validity but high LD decay bias. The MTAG eduPGS were PGS computed using MTAG. This is a method for analyzing statistics from genome-wide association study (GWAS) on different but genetically correlated traits (e.g., education and intelligence). Third, we used the MTAG eduPGS lead SNPs to construct an eduPGS with moderate within-population validity and LD decay bias (*N* = 1558 overlapping variants). The lead SNPs are the ones which had the most-significant association with education/intelligence within each cluster of LD-associated SNPs. The reasoning is that because SNPs with lower *p*-values (i.e., greater statistical significance) are more likely to be causal (Spencer et al., 2014), and because causal SNPs are more likely to be transethnically valid (Wang & Teo, 2015; Zanetti & Weale, 2018), using SNPs with a higher significance may increase the transethnic validity of eduPGS (Piffer, under review). Fourth, we used Lee et al.'s (2018) putatively causal variants to construct an eduPGS with minimal LD decay bias and low within-population validity (*N* = 111 overlapping variants).

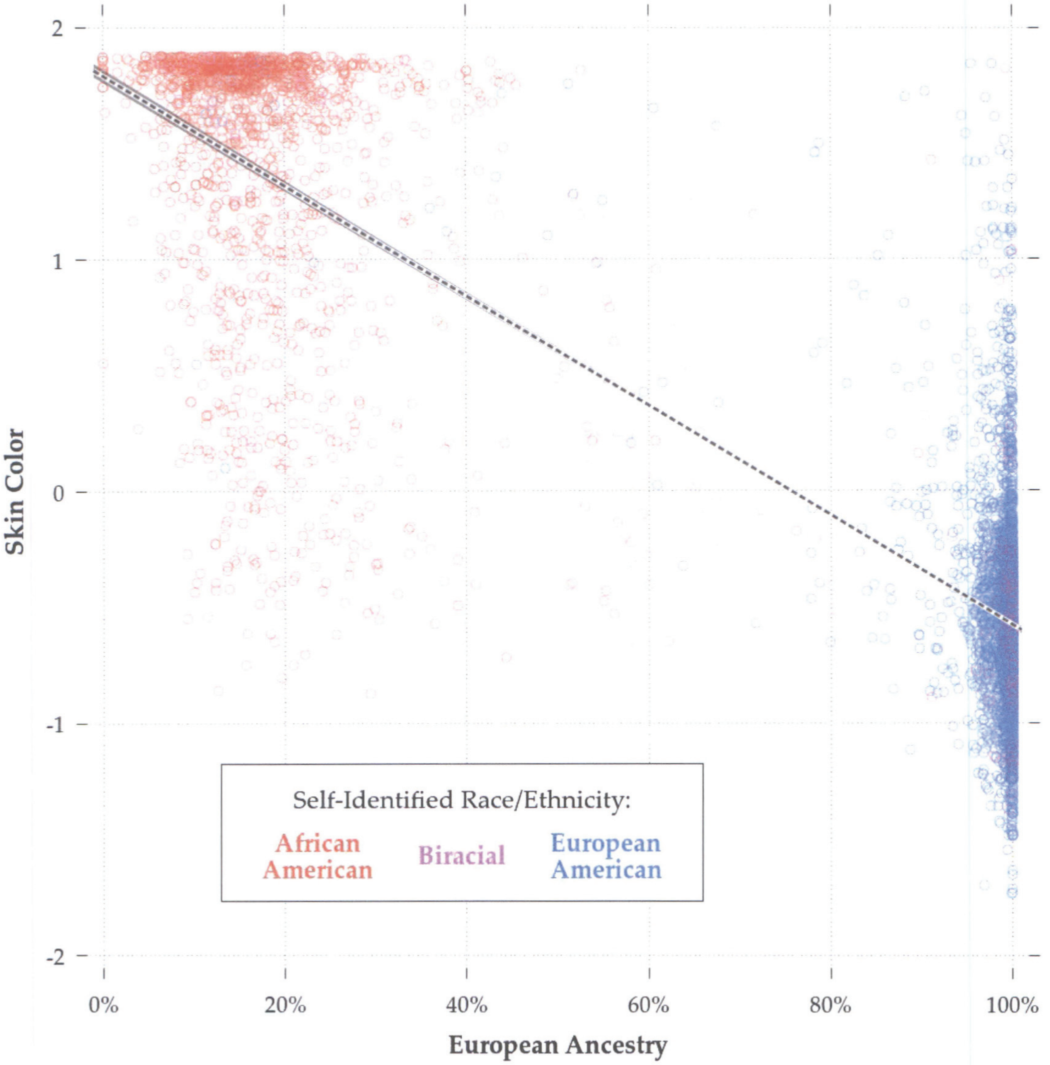


Figure 2. Regression plot of the relation between color (with higher values indicating darker color) and European genetic ancestry

3. Results

3.1. Descriptive Statistics

Table 2 shows descriptive statistics for all groups. AA|EAs were primarily European (80%) in ancestry. Correspondingly, they were closer to the EA means on SES, *g*, and color. The AA/EA *g* difference was $d = 0.984$ ($IQ_{EA} = 100$ versus $IQ_{AA} = 85.28$, with the biracial group scoring in-between at 97.96), while the SES difference was $d = 0.99$ and the difference in color was $d = 3.60$.

Table 2. Sample characteristics for the African, European, and biracial African-European-American participants

	Age (*SD*; *N*)	**% European (*SD*; *N*)**	**Mean SES (*SD*; *N*)**	**Mean *g* Score (*SD*; *N*)**	**Mean Color (*SD*; *N*)**
African	14.08 (3.75; 2,227)	0.187 (0.117; 2,228)	−0.57 (0.77; 2,180)	−1.01 (1.07; 2,179)	30.96 (5.87, 1,557)
Biracial African-European	13.15 (3.58; 232)	0.796 (0.289; 232)	0.16 (0.95; 230)	−0.14 (1.05; 228)	19.18 (7.61, 166)
European	13.76 (3.64; 4,937)	0.986 (0.059; 4,939)	0.33 (0.96; 4,909)	0.00 (1.01; 4,914)	14.70 (3.84; 3,862)

Note: Standard deviations and sample sizes appear in parentheses. SES = socioeconomic status.

3.2. Bivariate Relationships Among the Variables

Correlations between *g*, parental SES, genetic ancestry, and SIRE group appear in Tables 3 and 4. Correlations are as expected based on the cognitive and SES differences between different SIRE groups and the strong correlation between genetic ancestry and SIRE. However, this bivariate analysis is not very informative as outcomes are confounded with non-genetic causes and ancestry variables are negatively correlated by design since individual ancestry must sum to one. The relationship between parental SES and cognitive ability is likely stronger here than is typically reported (see, Sirin, 2005; $r = 0.29$) because we used both parents' education and a latent cognitive ability measure which reduces error that would normally attenuate the strength of this relationship.

The correlations between cognitive ability and both color and African ancestry are highest in the combined African-, African-European-, and European-American sample, intermediate in the combined mono- and bi-racial African-American sample, and lowest in the monoracial African-American sample. The difference in the magnitudes of the correlations is a result of the first sample having the most variability in ancestry and the last sample having the least. This difference is mostly a statistical artifact of restriction of range, which attenuates correlations (Hunter & Schmidt, 2004).

Table 3. Pairwise correlations among African-, African-European, and European-Americans

	Cognitive Ability	**SES**	**Euro. Ancestry**	**Afr. Ancestry**	**SIRE EA**	**SIRE AA**	**Color**	**EduPGS**
Cognitive Ability	1							
SES	0.406 (7,253)	1						
European Ancestry	0.411 (7,321)	0.412 (7,319)	1					
African Ancestry	−0.411 (7,321)	−0.412 (7,319)	−1.000 (7,399)	1				
SIRE EA	0.408 (7,321)	0.413 (7,319)	0.964 (7,399)	−0.964 (7,399)	1			
SIRE AA	−0.387 (7,321)	−0.395 (7,319)	−0.928 (7,399)	0.928 (7,399)	-0.930 (7,399)	1		
Color	−0.359 (5,534)	−0.354 (5,530)	−0.875 (5,585)	0.875 (5,585)	−0.838 (5,585)	0.812 (5,585)	1	
EduPGS	0.402 (7,321)	0.445 (7,319)	0.672 (7,399)	−0.672 (7,399)	0.645 (7,399)	−0.630 (7,399)	−0.614 (7,399)	1

Note: All values significant at $p < 0.0001$. Pairwise *N* in parentheses. SES = socioeconomic status, SIRE = self-identified race/ethnicity, eduPGS = education polygenic score.

Table 4. Pairwise correlations among self-identified monoracial African-American (below the diagonal) and self-identified monoracial and biracial African-European-American (above the diagonal)

	Cognitive Ability	**SES**	**Euro. Ancestry**	**Afro. Ancestry**	**Color**	**EduPGS**
Cognitive Ability	1	0.315* (2367)	0.251* (2407)	−0.251* (2407)	−0.188* (1690)	0.212* (2407)
SES	0.256* (2,140)	1	0.276* (2410)	−0.276* (2410)	−0.184* (1690)	0.244* (2410)
European Ancestry	0.086* (2,179)	0.054* (2,180)	1	−1.000* (2460)	−0.639* (1723)	0.522* (2460)
African Ancestry	−0.086* (2,179)	−0.054* (2,180)	−1.000* (1,854)	1	0.639*(1723)	−0.522* (2460)
Color	−0.084* (1,526)	−0.047 (1,524)	−0.389* (1,557)	0.389* (1,557)	1	−0.394* (1723)
EduPGS	0.112* (2,179)	0.119* (2,180)	0.328* (2,228)	−0.328* (2,228)	−0.187* (1,557)	1

Note: *Significant at $p < 0.01$. Pairwise *N* in parentheses. SES = socioeconomic status, eduPGS = education polygenic score.

3.3. Regression Analyses

Our primary focus is whether genetic ancestry predicts cognitive ability. The rationale for the analysis has been explained in detail by Kirkegaard et al. (2019). In order to determine if this is the case, we constructed a series of regression models. These models were applied to the African-

American and the combined African-, European-, and biracial African-European-American sample. Restricting the first analysis to SIRE African-Americans approximates the methodology used by Scarr et al. (1997). Our initial analysis used European ancestry (EUR) as the independent variable. We then added skin color to test discrimination (i.e., colorism) and other appearance-based hypotheses. Finally, we added SES to assess the effect of familial environmental influences on the observed gap, noting again that these measures are genetically confounded.

To provide results that can be easily compared across groups and studies, we standardize all values except European ancestry and then report the unstandardized beta coefficients. Doing so leaves only ancestry unstandardized. This allows one to interpret the beta coefficient for ancestry as a change in a standard deviation of cognitive ability over a change of 100% European ancestry. Following Kirkegaard et al. (2019), we do not report standardized betas (except in Table 5 for illustration), because, in the context of the effect of ancestry on outcomes, these are difficult to compare across groups. This difficulty arises because groups differ substantially in their variability in ancestry and its effects. As such, a change of one standard deviation of ancestry from one group has a substantially different meaning than the same from another group.

Table 5. Regression analysis for European ancestry as a predictor of *g* among monoracial African-Americans with controls for skin color (Model 2), and SES (Model 3) added

	Model 1a			**Model 1b**		
Predictor	*B*	*SE*	β	*B*	*SE*	β
Intercept	−1.16	0.04	−1.01	−0.82	0.06	−0.997
EUR	0.78***	0.19	0.09***			
Skin Color				−0.13***	0.04	−0.09***
SES						
Adjusted R^2	0.007			0.006		
N	2,179			1,526		

	Model 2			**Model 3**		
Predictor	*B*	*SE*	β	*B*	*SE*	β
Intercept	−1.05	0.09	−0.996	−0.83	0.09	−0.995
EUR	0.83***	0.24	0.10***	0.67**	0.23	0.08**
Skin Color	−0.07	0.04	−0.05	−0.06	0.04	−0.04
SES				0.36***	0.03	0.28***
Adjusted R^2	0.013			0.082		
N	1,526			1,500		

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Model 1b shows the results with color as an alternative predictor. EUR = European ancestry. SES = socioeconomic status.

In models that only included monoracial African-Americans, we found that European ancestry was always strongly and significantly related to cognitive ability. Skin color (assessed genetically with the highly accurate predictor (Chaitanya et al., 2018; Walsch et al., 2014) was associated with cognitive ability (Model 1b, Table 5), but made no significant incremental contribution when ancestry was also in the model (Model 2, Table 5). Results could still be due to phenotypic confounding from other appearance variables. To test this possibility, we fitted a number of models including skin, hair, and eye color. We found that none of these features had significant effects on their own, except for brown eye color, which was positively related to cognitive ability, but with a large standard error. These results are shown in the R notebook. The last monoracial African-American model (Model 3, Table 5) included SES, which had a considerable effect on cognitive ability. However, the so-called sociologist's fallacy (Jensen, 1998; Sesardic, 2005) may be at play here. That is, controlling for parental SES also controls for genetic effects on SES which may be shared with cognitive ability.

As an alternative to the model with only monoracial African-Americans, we next included both individuals who identified as monoracial African- and biracial African-European-American. This is equivalent to the "broadly-defined" African-American group used by Kirkegaard et al. (2019). These results are shown in Table 6. In Model 2, with self-identified European-American race (which, in this case, indicates biracial status) as a covariate, the beta for European ancestry ($B_{EUR} = 0.85$; $N = 2407$) was comparable although slightly higher than that for the monoracial African-American sample ($B_{EUR} = 0.78$; $N = 2179$).

Table 6. Regression analysis for European ancestry as a predictor of *g* among monoracial and biracial African-European-Americans with controls for self-identified race/ethnicity (Model 2), skin color (Model 3b), and socioeconomic status (Model 4) added

	Model 1		**Model 2**		**Model 3a**		**Model 3b**		**Model 4**	
Predictor	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Intercept	−1.22	0.03	−1.17	0.04	−0.62	0.05	−1.07	0.08	−0.83	0.08
EUR	−1.21***	0.09	0.85***	0.15			0.80***	0.19	0.55**	0.19
SIRE: EA			0.35**	0.12			0.27*	0.13	0.18	0.13
Skin Color					-0.25***	0.03	−0.06	0.04	−0.05	0.04
SES									0.37	0.03
Adjusted R^2	0.063		0.066		0.035		0.065		0.134	
N	2,407		2,407		1,690		1,690		1,664	

Note: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. EUR = European ancestry. SES = socioeconomic status, SIRE = self-identified race/ethnicity.

Comparing Model 2 and Model 3b, biracial status had a significant effect independent of European ancestry. This could be due to selective patterns of self-identification, whereby choosing to report that one is a European-American, regardless of one's genetic ancestry, is associated with greater cognitive ability. This appears to have been the case as evidenced by Figure 3, which shows the relationship between European ancestry and cognitive ability. Being above the regression line was associated with a greater likelihood of biracial African-European self-identification regardless of one's actual ancestry. Whatever the case, unlike with self-reported European-American race, skin color did not have significant predictive validity independent of European ancestry. The comparison between Model 3a and Model 3b shows that the inclusion of ancestry reduced the effect of color from highly significant to non-significant ($B_{color} = -0.25$ in 3a and $B_{color} = -0.06$ in 3b, $N = 1,690$). As in the models for monoracial AAs, the inclusion of parental SES reduced, but did not eliminate, the effect of European ancestry (Model 4).

The next set of models utilized the whole sample of African, European, and biracial African-European-Americans. These results appear in Table 7. The first model revealed a substantial impact of European ancestry on cognitive ability. The next model found the same, but also produced a significant effect of SIRE European-American (but not SIRE African-American) self-identification.

That is, some incremental validity existed which was associated with self-identification as European-American and this did not occur for self-identification as African-American. The next model included physical appearance in the form of skin color. As with the African-American-only and the combined African- and biracial African-European-American sample, this variable had no significant effect on cognitive ability in the full sample with ancestry included (Model 3b). Our final model included SES. The inclusion of SES reduced the effect of SIRE European-American, although it still remained significant. Nonetheless, in this model, European ancestry was the dominant determinant of cognitive ability. The effect of SES on cognitive ability was essentially the same in the full sample as it was in the monoracial African-American model. The fact that the effect of SIRE European-American weakened with the inclusion of SES is perhaps notable, as it may imply that selective self-identification as European-American is driven in part by social processes. It is clear here though that self-identification and other-identifiable physical characteristics are much less important as determinants of cognitive ability than is genetic ancestry.

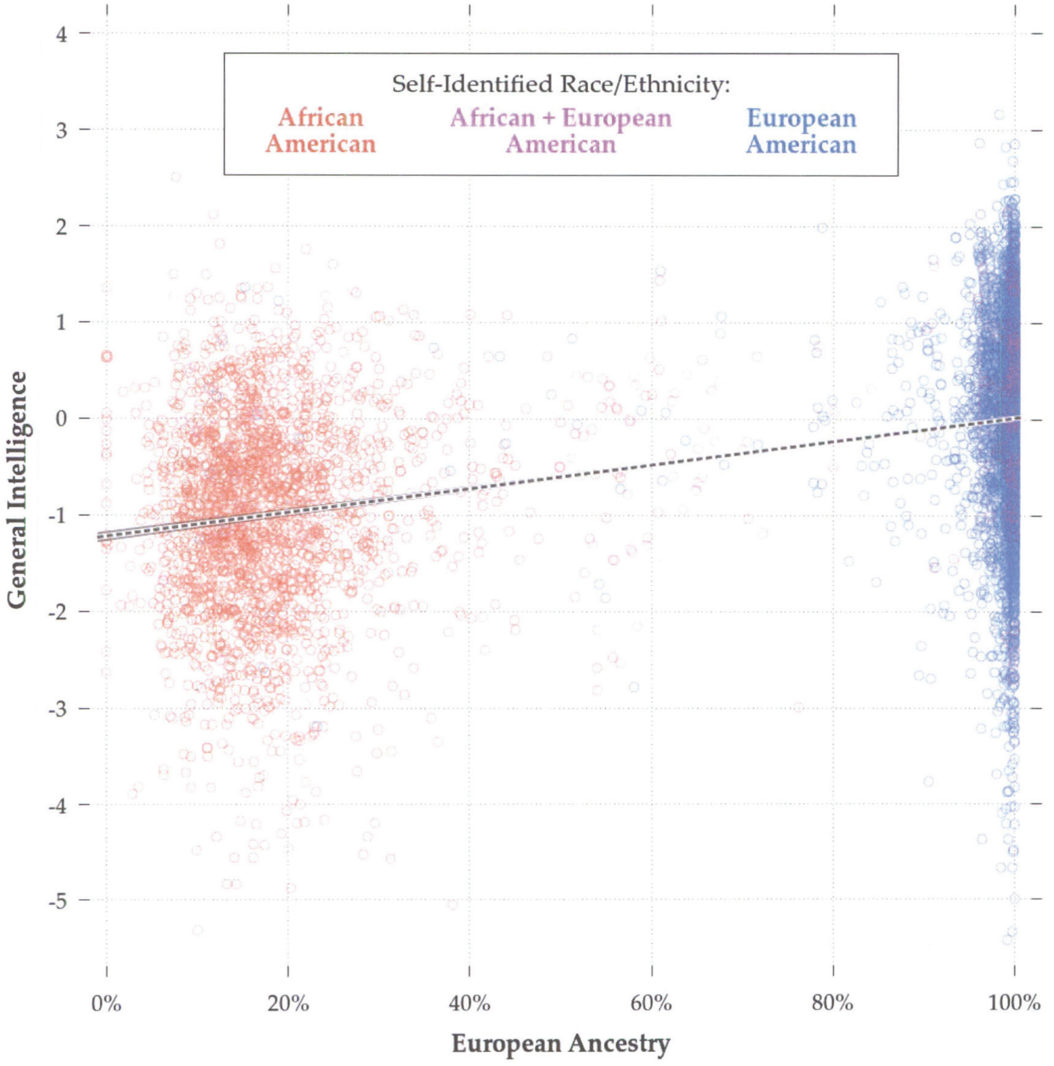


Figure 3. Regression plot for the relation between *g* and European ancestry ($r = 0.411$)

Table 7. Regression analysis for European ancestry as a predictor of *g* among African-, European-, and biracial African-European-Americans with controls for SIRE (Model 2) skin color (Model 3), and SES (Model 4) added

	Model 1		**Model 2**		**Model 3a**		**Model 3b**		**Model 4**	
Predictor	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Intercept	−1.22	0.03	−1.15	0.09	−0.28	0.01	−1.09	0.11	−0.91	0.11
EUR	1.23***	0.03	0.75***	0.13			0.74***	0.16	0.59***	0.15
SIRE: EA			0.41***	0.11			0.35**	0.12	0.22*	0.11
SIRE: AA			0.00	0.08			−0.02	0.09	0.01	0.08
Skin Color					-0.40***	0.01	−0.02	0.03	−0.01	0.03
SES									0.34***	0.01
Adjusted R^2	0.169		0.170		0.129		0.166		0.238	
N	7,321		7,321		5,534		5,534		5,488	

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. EUR = European ancestry. SES = socioeconomic status, SIRE = self-identified race/ethnicity.

Cumulatively, European ancestry was consistently positively related to cognitive ability in all models. SES was similarly related to cognitive ability. European-American SIRE also had a significant effect independently of genetic ancestry, indicating that differences were not wholly explained by ancestry. Skin color was not significantly related to cognitive ability in any of our models which included genetic ancestry. However, it was related to cognitive ability in models not including ancestry, as found in previous studies (Fuerst et al., 2019; Hu et al., 2019; Kirkegaard & Fuerst, 2017; Lynn, 2002) (for a review of older studies, see Shuey, 1966).

Recent immigrants from Africa in the sample may have affected the scores reported here as migrants may have been selected for cognitive ability when they emigrated from their home countries. As such, we conducted a robustness test by rerunning the analyses excluding individuals with one percent or less European ancestry. This exclusion did not substantially affect the results (reported in the R notebook, Model 17b). Thus, confounding due to migration likely did not influence our findings.

3.4. Cognitive Ability and Educational-Related Polygenic Scores (eduPGS)

We examined the validity of four cognitive ability and education-related PGSs reported by Lee et al. (2018). The eduPGS were derived from European populations. First, we used the eduPGS based on all variants trained without the 23andme cohort (*N* = 7,762,369 variants overlapping in the present dataset). The goal was to construct a predictor with good within-discovery population (European, in this case) validity, but very high LD decay bias, and thus low transethnic validity. Second, we used MTAG eduPGS with 10k variants (*N* = 8442 overlapping variants) to construct a predictor with good within-discovery population validity and reduced LD decay bias. Third, we used the MTAG eduPGS lead SNPs (*N* = 1558 overlapping variants) to see if we could reduce LD decay bias and increase transethnic validity while retaining within-discovery population validity. Finally, we used the putatively causal variants (*N* = 111 overlapping variants in our dataset) to construct an eduPGS with minimal LD decay bias, but low within-discovery population validity. At some point, screening SNPs based on *p*-values may lead to lower validity in both the discovery population and other-ethnic populations, despite having relatively high transethnic validity (Piffer, under review). The goal is to identify PGS which have both high within-discovery population and transethnic validities.

The correlations with g are shown in Table 8, with results for European-Americans above the diagonal and results for (monoracial) African-Americans below. Since the PNC is independent from the samples analyzed by Lee et al. (2018), the results for European-Americans constitute an independent replication of those reported by Lee et al. (2018). This may be of interest to some because our dependent variable was a good measure of g.

Of the PGS, the MTAG_10k PGS had the highest validity in both the European-American and African-American samples, although the validity for African-Americans was reduced by 51%. The results for MTAG_10k and MTAG eduPGS lead SNPs correlated highly with one another ($r > 0.80$) in both the African- and European-American samples. These eduPGS also had similar relationships with cognitive ability.

As expected, the GWAS eduPGS had high within-source population validity but low transethnic validity. While significant in the African-American sample, the validity was only 0.044/0.225 = 20% of that in the European-American sample. This low transethnic validity occurred because the 7,762,369 variants in conjunction with high LD decay produced considerable noise. Due to being based on only 111 overlapping variants, the putatively causal eduPGS had low validity within both samples but had the highest relative transethnic validity: 0.031/0.058 = 54%.

Table 8. Pairwise correlations between cognitive ability and education/intelligence-related polygenic scores (European-American above the diagonal, African-American below)

	Cognitive Ability	**Putative Causal**	**GWAS_edu PGS**	**MTAG_10K_eduPGS**	**MTAG_Lead eduPGS**
Cognitive Ability	1	0.058*** (4,914)	0.225*** (4,914)	0.227*** (4,914)	0.210*** (4,914)
Putative Causal_edu PGS	0.031 (2,179)	1	0.216*** (4,939)	0.315*** (4,939)	0.348*** (4,939)
GWAS_edu PGS	0.044* (2,179)	0.135*** (2,228)	1	0.645*** (4,939)	0.574*** (4,939)
MTAG_10K_eduPGS	0.112*** (2,179)	0.228*** (2,228)	0.484*** (2,228)	1	0.837*** (4,939)
MTAG_Lead_PGS	0.094*** (2,179)	0.266*** (2,228)	0.453*** (2,228)	0.801*** (2,228)	1

Note: *Significant at $p < 0.05$; **Significant at $p < 0.01$; ***Significant at $p < 0.001$. N in parentheses. GWAS = genome-wide association study (standard one-at-a-time regression), MTAG = multi-trait analysis of GWAS, eduPGS = education polygenic score, lead = genome wide-significant hits only.

Consistent with theory (Piffer, under review), the eduPGS with the highest validity in both the African- and European-American samples was the MTAG_EA_10K ($r = 0.112$ & $r = 0.227$, respectively). As such, in the following analyses, we restricted consideration to these scores. While retaining some validity in the African-American sample, the MTAG_10k eduPGS showcased predictive bias. Indeed, a Fisher r-to-z Test showed that the differences in the correlations for African and European-Americans was highly significant ($z = -4.6$, $p < 0.01$). This difference is illustrated in Figure 4, which shows the regression plot for eduPGS and g for European (red) and African-Americans (blue).

The slope of the regression line and the intercepts for the African-American sample are both lower. As Zanetti & Weale (2018) note, differences in SNP predictivity could be due to (1) lack of statistical power, (2) allele frequency differences, (3) linkage disequilibrium, or (4) differences in true causal variant effect sizes/directions. These authors suggest that the last explanation is generally not necessary. Regardless, a combination of these factors could lead to PGS bias in favor of Europeans (or Africans). There are a number of possible ways in which such bias could have arisen. For example, imagine the case for more recent polymorphisms. These more recent polymorphisms will probably have a higher frequency in the discovery population. If recent selection favored IQ-promoting variants and purged IQ-reducing variants, the more recent polymorphisms will be IQ-promoting, which will result in bias in favor of the discovery population. There is some evidence of an increase in PGS over time (Woodley et al., 2017). Of course, the reverse would be true, if recent selection disfavored high IQ alleles. However, given that most mutations are deleterious (Uricchio et al., 2019), if the PGS contain a large number of recent variants it is likely that they'll be biased against their discovery populations. For disease phenotypes, Kim, Patel, Teng, Berens, & Lachance (2018) showed how this bias can function and found that correcting for differences in derived vs. ancestral polymorphisms can improve the transracial predictivity of PGS. In a supplementary analysis, conducted to address a reviewer's point, we did not find evidence of a substantive (or even statistically significant) effect of young vs. old polymorphisms. Thus, the concern raised regarding bias due to polymorphism age (e.g., Kim et al., 2018) does not seem to be a problem in this case and so we did not create polymorphism age-corrected PGS.

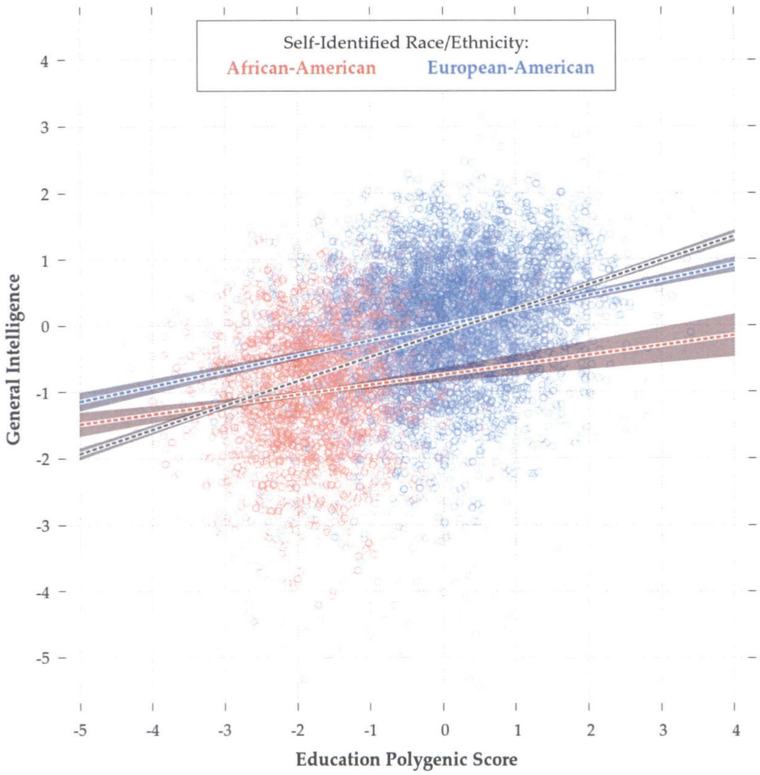


Figure 4. Regression plot for the predictive validity of MTAG 10k eduPGS with respect to *g* in the African-American (Red; $r = 0.112$) and European-American (Blue; $r = 0.227$) samples

Table 9 shows eduPGS means and standard deviations for African, biracial African-European, and European-Americans. The eduPGS difference is large at $d = 1.89$ and it is larger than the phenotypic gap in *g*. This would be expected even if the PGS were unbiased given the imperfect relationship between PGS and *g*. Beyond this and possible bias in the PGS, another possibility is that the PGS differences between groups are primarily in common variants, while the differences between individuals within groups are in both common variants and rare variants that arose recently by mutations and which are under mutation-selection balance. Given this scenario, PGS which indexed mostly common variants between populations would manifest as reduced phenotypic differences even assuming consistent validity.

Table 9. Mean racial differences in education polygenic score (MTAG_10k)

	N	*M*	*SD*
European-American	4939	0.00	1.00
Biracial African-European-American	232	−0.59	1.15
African-American	2228	−1.79	0.81

The association between eduPGS (MTAG_10k) and *g* in the African-American sample could, in principle, result from confounding with environmental factors. This could be an indirect effect of the association between *g* and European ancestry or, alternatively, color. Alternatively, the variants in the eduPGS could be "chopstick genes" (Hamer & Sirota, 2000) which have a statistical but no causal effect. If eduPGS were merely indexing European ancestry, one would expect the GWAS eduPGS (which is based on more than 7 million additional variants) to exhibit an association as high as the one we found with MTAG_EA_10k eduPGS. This is because the GWAS_eduPGS, with its greater number of variants, is likely to be a better index of ancestry. This does not appear to be the case. Regardless, to explore this possibility further, we ran regression models with European ancestry (Model 1b) and color scores (Model 2b) for the monoracial African-American and combined African-, European-, and biracial African-European-American samples. The results are shown in Tables 10 and 11 respectively.

Table 10. Regression results for the effect of education polygenic score on cognitive ability among monoracial African-Americans

	Model 1b		**Model 2b**	
Predictor	***B***	***SE***	***B***	***SE***
Intercept	−0.88	0.08	−0.75	0.12
EUR	0.50*	0.21	0.51*	0.25
eduPGS	0.12***	0.03	0.13***	0.04
Color			−0.07	0.04
Adjusted R^2	0.014		0.021	
N	2,179		1,526	

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. EUR = European ancestry. eduPGS = education polygenic score.

Adjusting for European ancestry does not reduce the validity of eduPGS in the African-American sample (Table 10; Model 1b), nor does controlling for skin color. In fact, the validity here increases slightly relative to that found among European-Americans. Table 12 shows the regression results for European-Americans (Model 2). Comparing the monoracial African-American SIRE beta (Model 1b, Table 10) to the European-American beta (Model 2, Table 12) gives an African-American validity 52% the size of the European-American validity.

Table 11. Regression results for the effect of education polygenic score on cognitive ability among African, biracial African-European, and European-Americans

	Model 1b		**Model 2b**	
Predictor	***B***	***SE***	***B***	***SE***
Intercept	−0.73	0.09	−0.64	0.12
EUR	0.29*	0.13	0.27	0.16
SIRE: EA	0.44***	0.10	0.37**	0.11
SIRE: AA	0.04	0.07	0.01	0.09
eduPGS	0.21***	0.01	0.20***	0.01
Color			−0.01	0.03
Adjusted R^2	0.200		0.196	
N	7,321		5,534	

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. EUR = European ancestry, SES = socioeconomic status, SIRE = self-identified race/ethnicity, eduPGS = education polygenic score.

Table 12. Regression results for the effect of education polygenic score on cognitive ability among European-Americans

	Model 1		**Model 2**		**Model 3**	
Predictor	***B***	***SE***	***B***	***SE***	***B***	***SE***
Intercept	−0.45	0.25	0.02	0.24	−0.03	0.27
EUR	0.45	0.25	−0.02	0.24	0.04	0.28
eduPGS			0.23***	0.01	0.22***	0.02
Color					0.02	0.04
Adjusted R^2	0.000		0.051		0.50	
N	4,914		4,914		3,844	

Note: *$p < 0.05$, **$p < 0.01$, **$p < 0.001$. EUR = European ancestry, SES = socioeconomic status, SIRE = self-identified race/ethnicity, eduPGS = education polygenic score.

If we take the beta of 0.124 in the AA-only sample (Model 1b, Table 10) as an estimate of the LD-unbiased effect of eduPGS on *g*, the *g*-score differences due to eduPGS would be 0.124 × 1.89 (i.e., the Cohen's *d* for the African-/European-American eduPGS difference) = 0.23. This implies that 20%–25% of the difference in *g* can be accounted for by presently known cognitive ability- and education-related PGS.

Concerns have been raised that the eduPGS computed by Lee et al. (2018) may be biased against non-Europeans due to the inclusion of population-specific variants (Thompson, 2019) (for a more general discussion see Kim et al., 2018). These population-specific variants would have very low frequency in non-European populations. Following Piffer's suggestion (Thompson, 2019), as a robustness check, we computed MTAG eduPGS excluding variants with MAF < 0.01 (leaving 7636 overlapping variants) and <0.05 (leaving 7172 overlapping variants) among African lineages. The 1000 Genomes reference samples were used to determine the African MAF. This exclusion had no substantive effect on the eduPGS validity with respect to *g* for either African or European-Americans. For SIRE monoracial African-Americans rMAF $\geq 0.01 = 0.112$ and rMAF $\geq 0.05 = 0.114$.

Theory predicts that, in the absence of countervailing confounds, variants which show concordant as opposed to discordant effects across racial groups, will exhibit higher validity in non-discovery populations (Piffer, personal communication, February 3, 2019; under review). This is because variants which show transracially concordant effects are more likely to be causal and also because causal variants, as opposed to variants in LD with causal ones, are more likely to have transethnic predictivity (Marigorta & Navarro, 2013; Wang & Teo, 2015). These predictions are borne out in this sample. Among European-Americans, the predictivity for concordant MTAG_10k eduPGS was higher (*r*concordant = 0.223 vs. *r*discordant = 0.145), and among African-Americans, it was even more so (*r*concordant = 0.223 vs. *r*discordant = -0.128). Similarly, when we grouped variants by the *p*-value for the interaction between race and the effect on *g* (with higher *p*-values signifying less discordance in effect), among African-Americans, the variants with higher *p*-values were substantially more predictive of *g* (*r*high_*p* = 0.141 vs. *r*low_*p* = 0.035), while among European-Americans, this was not the case (*r*high_*p* = 0.265 vs. *r*low_*p* = 0.263). Cross-validation confirmed that the concordant PGS had higher predictive validity among the non-discovery population. However, the concordant PGS subset did not seem to have higher predictivity, based on cross-validation, than the combined set. The identical validity for concordant PGS could not be sustained in cross-validation. Thus, we do not use these disaggregated PGS in further analysis.

3.5. Path Analysis of Cognitive Ability and Educational-Related Polgenic Scores

Finally, we depict our results in path model form. As seen in Figure 4, substantial differences exist in the transethnic validity of the eduPGS. We therefore limited our path analysis to monoracial African-Americans. The model was run with the lavaan R package with missing data handled using listwise deletion. (Missing color scores were not MCAR, so we did not impute data.) Results are shown in Figure 5. We leave European ancestry unstandardized, as in previous analyses, while the other variables were standardized. Thus, the total effect of European ancestry on cognitive ability is calculated as (1) the unstandardized path coefficient between European ancestry and cognitive ability (B = 0.512) plus (2) the unstandardized path between European ancestry and color times the standardized path coefficient between color and cognitive ability (−2.209 × −0.066 = 0.146) plus (3) the unstandardized path between European ancestry and eduPGS multiplied by the standardized path coefficient between eduPGS and cognitive ability (2.515 × 0.131 = 0.329).

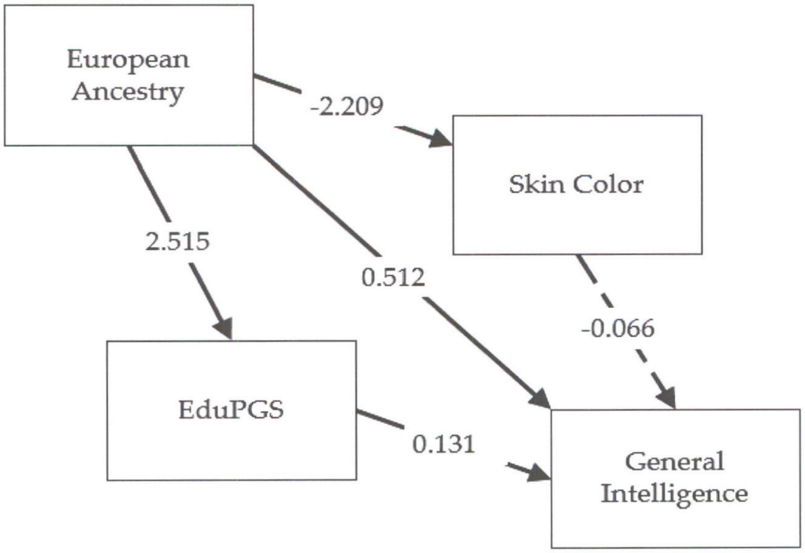


Figure 5. Path diagram for relation between European Ancestry, skin color, education polygenic score, and *g* in the African-American sample. eduPGS = education polygenic score (*N* = 1526)

Table 13 shows the estimates for the different paths. From the table, while European ancestry and eduPGS had significant effects, color did not (although it was trending toward significance, $p = 0.12$). Additionally, the covariance between color and eduPGS was not significant. The association between European ancestry and *g* is more substantially mediated by the cognition-related eduPGS than by color scores. An analysis of the effect of color within sibling pairs is included in the supplement but the results are extremely tentative due to a lack of statistical power and the unconfirmed validity of our measure of Skin Color.

Table 13. Detailed results for path diagram

			Unstandardized Estimate	***SE***	***P*-value**	**Lower 95% CI**	**Upper 95% CI**	**Standardized Estimate**
EUR	→	*g*	0.512	0.255	0.045	0.013	1.012	0.059
EUR	→	eduPGS	2.515	0.156	0.000	2.209	2.821	0.381
eduPGS	→	*g*	0.131	0.036	0.000	0.060	0.202	0.099
Skin Color	→	*g*	-0.066	0.043	0.120	-0.150	0.017	0.043
EUR	→	Skin Color	-2.209	0.133	0.000	-2.469	-1.949	-0.392
Skin Color	~	eduPGS	-0.020	0.012	0.089	-0.044	0.003	-0.044

Note: EUR = European ancestry, eduPGS = education polygenic score.

3.6. The Spearman–Jensen Hypothesis

Spearman's hypothesis holds that racial differences in cognitive ability are largely the result of differences in *g*. This theory predicts a strong positive correlation between subtest *g* loadings and the vector of group differences. We test this theory using Jensen's method of correlated vectors (MCV; Jensen, 1998) by examining the correlation between the vector of *g* loadings, the vector of subtest heritability, and four vectors of group differences. The vector of *g* loadings is the *g* loadings for the 10 subtests found to be MI for African and European-Americans. The vector of heritability is the average of the African- and European-American heritabilities reported by Mollon et al. (2018). The vectors of group differences were: (1) the correlation with European ancestry (for the combined sample), (2) the correlation with MTAG 10K eduPGS for the European-American sample, (3) the correlation with MTAG 10k eduPGS for the monoracial African-American sample, and (4) the African-/European-American mean group differences. These results are shown in Table 14.

Spearman's hypothesis is strongly confirmed in our data. Not only is there a large Jensen effect on European ancestry, there is an equally large Jensen effect for the subtest correlations with the MTAG eduPGS. The correlations with the heritabilities are also positive, although these were estimated based on small samples (as far as behavior genetics goes) and thus this vector correlation is downwardly biased. If one includes the remaining 5 subtests, the correlations are about the same magnitude.

Table 14. Results from Jensen's Method of Correlated Vectors

	***g* Loading**	**B-W h^2**	**Ancestry *r***	**eduPGS *r* (European-American)**	**eduPGS *r* (African-American)**	**African/ European Gap**
g loading	1					
B-W h^2	0.372	1				
Ancestry *r*	0.941	0.467	1			
eduPGS *r* (European-American)	0.892	0.604	0.993	1		
eduPGS *r* (African-American)	0.901	0.600	0.836	0.846	1	
African/European gap	0.929	0.478	0.997	0.926	0.813	1

Note: *N* (subtest) = 10, eduPGS = education polygenic score, h^2 = heritability.

3.7. Assessment of Measure Invariance with Respect to Ancestry

Initially, we confirmed that MI held between self-identified African- and European-Americans. However, it is also possible to examine if MI holds with respect to genetic ancestry using LSEM (see Hildebrandt et al., 2016). A finding of MI along the whole range of admixture would preclude many group-specific factors which are commonly proposed to explain potential admixture-ability or other admixture-outcome relationships.

We fit an LSEM on the full African- and European-American sample (*N* = 7399) using 21 focal points (i.e., 5% ancestry increments ranging from 0% to 100%). From 0% to 100% European ancestry there was no deviation from MI. While the mean levels in *g* and complex reasoning increased linearly, the mean levels in episodic memory and executive functioning decreased linearly. MI appears to be tenable across the whole range of ancestry and, as well, the relationship between latent abilities and genetic ancestry comports with Spearman's hypothesis.

We additionally fitted LSEMs with SES. When standardized SES was modeled from its lowest to its greatest value, it showed measurement invariance but failed to account for the differences in the latent means between racial groups despite having a similar, albeit smaller, effect compared to European admixture when modeled as a background variable in the MGCFA model. The important takeaways from the LSEM results are that MI with respect to both SES and admixture is tenable, admixture can fully account for group mean differences in latent abilities while our present measure of SES cannot, and, as MI implies that the differences between groups on tests are due to differences in means and variances alone, this appears to be the case with admixture and race generally.

4. Discussion

We conducted an admixture study to assess the association between cognitive ability and European ancestry in the African- and European-American populations. Earlier researchers (e.g., Loehlin et al., 1973; Scarr et al., 1997) were unable to detect significant relationships between ancestry and cognitive ability. We, however, found a strong, significant association between European genetic ancestry and cognitive ability, confirming the recent results of both Kirkegaard et al. (2019) and Warne (2019).

4.1. Global Admixture Analysis and Past Research Findings

The betas we calculated for European ancestry were very similar across models: B = 0.78 for the self-identified monoracial African-American-only sample (Model 1, Table 5) and B = 0.75 for the combined African- and European-American sample (with controls for SIRE; Model 2, Table 6). The former beta corresponds to a bivariate correlation of only $r = 0.086$ ($N = 2179$; Table 4). This small correlation is misleading as to its importance. The correlation is small in magnitude simply because the variation in European ancestry in the sample is low (or restricted). As Jensen (1981) noted, constrained variability results in attenuated admixture correlations. Because of this, a small correlation within a population with restricted ancestry variance can be consistent with a large overall effect of admixture. For example, our African-American sample averaged 18.7% (*SD* = 11.7%) European ancestry. Compared with European-Americans (98.6%), this is a difference of (98.6%–18.7%)/11.7%= 6.83 standard deviations of European ancestry. Thus, the predicted difference in cognitive ability between African- and European-Americans (or, rather, African-Americans with the same amount of European ancestry as European-Americans in this sample), based on our observed bivariate association, would be $r = 0.086 \times 6.83 = 0.587$ SD. This is relative to the African-/European-American difference of $d = 0.984$ in cognitive ability (from Table 1). Thus, this small $r = 0.086$ correlation is consistent with 60% of the mean difference being explained by genetic ancestry.

Note that the correlation Scarr et al. (1997) found was 0.05/0.086 = 58% of the correlation we found. The attenuation in correlation is likely attributable to the reduced validity of their admixture index and the lower heritability of their measure of cognitive ability. Thus, our study largely vindicates Jensen's (1981) reanalysis of the Scarr et al. (1997) data.

We found very large Jensen effects for both the African-/European-American difference ($r = 0.929$) and the association between ancestry and *g* loadings ($r = 0.941$). The finding of a Jensen effect on ancestry is also consistent with the results presented by Scarr et al. (1997). As Loehlin (2000) seems to have implied, the association between admixture and ability in the Scarr et al. (1997) sample comported with the typical weak form of Spearman's hypothesis in that European ancestry was positively related ($r = 0.1$) with Raven's Progressive Matrices scores and negatively related ($r = -0.12$) with a memory measure. Indeed, using average *g* loadings provided by Jensen (1985) (cf. Scarr, 1981) the correlation between the African-/European-American score differences and *g* loadings is $r = 0.64$, while the correlation between the African-American-only *g* loadings and the score differences between the most- and least-African in this group is $r = 0.84$.

The only other study which has examined Spearman's hypothesis in the context of intra-SIRE ancestry-related differences is Hu et al. (2019). The authors studied participants from the National Longitudinal Survey of Youth 1997. Among African-Americans, Hu et al. (2019) found that darker color was associated with lower *g* scores while parent-reported European ancestry was associated with higher *g* scores. Jensen effects were found for both indices of ancestry. It seems that Spearman's hypothesis generalizes to ancestry-related differences within self-identified racial groups.

Our study goes beyond those of Kirkegaard et al. (2019) and Warne (2019) in that we (as with Hu et al., 2019) established both SFI and the weak form of Spearman's hypothesis regarding the African-/European-American difference. On the basis of this result we conclude that the SIRE differences in this battery are typically-sized and consistent with prior investigations of ability patterns in African- and European-American samples. Because the test battery is unbiased and it assesses the same traits in both populations, the difference observed here is a real difference in the cognitive abilities being measured.

We go beyond Hu et al. (2019) in that we included a measure of both color and genetic ancestry in our regression analyses. As predicted by the human capital model (Hu et al. (2019), Prediction 7), we found that color had no incremental validity beyond ancestry. This lack of incremental validity for color is possibly due to power. Color scores were available for only a subset of individuals since several of the arrays did not include all the SNPs needed by the algorithm to compute the scores. Thus, we cannot completely rule out an incremental effect of color in the present dataset. We can, however, conclude that the association between European ancestry and *g* is not secondary to that between color and *g*.

It is worth noting that in the path model we did not find a significant association between skin color and eduPGS. This could be due to a lack of statistical power. This may, however, suggest a limited role for assortative mating (or, for that matter, pleiotropy), which has been suggested by some (e.g., Jensen, 1998) in explaining the color-IQ association among African-Americans.

Beyond color, we also controlled for SES. Although not as strong of a predictor as ancestry, SES was a useful predictor of cognitive ability which modestly attenuated the effect of ancestry on cognitive ability. These results, however, are expected in both genetic and non-genetic models because environmental variables and familial characteristics like SES also reflect genetic influences (Jensen, 1998; Plomin & Bergeman, 1991; Vinkhuyzen et al., 2010). Controlling for these variables may therefore spuriously weaken the overall association between ancestry and the relevant genetic factors that actually explain their influence. Conditions in which cross-sectional data reflect real mediation are exceptional (Cole & Maxwell, 2003) so future investigations should attempt to tease out causality with longitudinal designs and local admixture mapping. A particularly strong research design would be to investigate sibling variation in admixture just as Visscher et al. (2006) did with relatedness. Very large samples are needed for this design if using full siblings; if data can be found using half- or adopted-siblings, then the required sample sizes would be more reasonable.

4.2. eduPGS Findings and Past Research

We also evaluated the transethnic validity of eduPGS. We found that the eduPGS with the highest validity for *g* in both the African- and European-American samples was the MTAG_EA_10K set. Although the validity in the African-American sample was approximately half of that in the European-American sample (rAA = 0.1115; rEA = 0.2269), the relations were statistically significant in both populations ($p < 0.0001$). As with Piffer (2019) we found large African-/European-American differences in these eduPGS ($d = 1.89$). Using the beta in the African-American sample and controlling for the effect of European ancestry (B = 0.124; Model 1b; Table 10), we estimate that the known eduPGS can naïvely explain as much as 20%–25% of the African-/European-American intelligence gap.

Removing variants with low MAF in African 1000 Genomes lineages had little effect on the validity of MTAG_10K eduPGS among African and European-Americans. Thus, contrary to some arguments, European-specific alleles do not seem to be biasing prediction with eduPGS. Our results corroborate those of Piffer (2019) who found a strong ecological correlation between MTAG-derived SNPs and population IQ ($r = 0.86$). We also found, using both regression and path analysis that, while the eduPGS mediates the association between European ancestry and cognitive ability, skin color scores do not.

4.3. Alternative Explanations and Methods to Further Reduce Uncertainty

At least one researcher has advanced the possibility that admixture designs are hopelessly confounded due to historical assortative mating (Centerwall, 1978). This theory supposes that the European ancestors of African-Americans were genetically disposed toward either lower or higher cognitive ability in a systematic way. Since Piffer (2019) found that other African descent groups have similarly low eduPGS, this explanation would necessitate that the eduPGS are not indexing genetic differences related to cognitive ability in African groups. Determining if this is the case is important for future research. We can directly investigate this hypothesis and further test the transethnic validity of the eduPGS with local admixture mapping. If the theory implied here is correct, European ancestry segments should have systematically higher or lower eduPGS in African-American samples. Relatedly, if eduPGS lack transethnic validity in Africans, they will not be associated with specific regions of the brain, biological variables, or consistent patterns of gene expression as they are in Europeans.

Our data are compatible with a between-group heritability (variance explained by European ancestry) of between 50% and 70% depending on the model chosen (see Scarr et al., 1997, p. 85). This estimate of between-group heritability is consistent with Rushton and Jensen's (2005) hereditarian model, according to which 50%–80% of the African-/European-American cognitive difference is due to genetic differences.

While the statistical mediation by PGS scores suggests that genetic factors may be involved, as discussed in detail by Kirkegaard et al. (2019), we cannot rule out many types of confounding environmental variables with this research design. Global admixture analysis results are suggestive and should only be considered a first step for investigating the effects of admixture on a trait. We suggest to attempt replication of the current results using a nationally representative sample and then, if these findings are confirmed, proceed to admixture mapping (local admixture analysis). This is the standard approach taken in medical epidemiology.

We suggest two approaches to further reduce the uncertainty regarding the causes of the African-/European-American cognitive ability gap. First, attempts should be made to replicate the current results using other samples (e.g., the Add Health study; Dolan, 2000). Second, local ancestry analysis/admixture mapping to examine the regions of the genome where the association with ancestry is most pronounced are a natural follow-on project. The rationale of such an analysis has been explained by others (e.g., Hu et al., 2019; Kirkegaard et al., 2019; Shriner, 2013; Winkler et al., 2010). It would also be worthwhile to attempt to replicate these results in admixed American populations outside of the United States (e.g., Brazil, Colombia, etc.). This research project can and should be expanded to other ethnic groups both in and outside of the United States. Examples include Aborigines in Australia, Cape Coloureds in South Africa, and the Mestizo population in Mexico. For some groups, it may be of utility to examine differences in broad abilities (e.g., spatial or mathematical ability) instead of general ability as done here, as differences may not be general.

4.4. General Conclusion

Rushton and Jensen (2005) called for modern genetic studies to test the hereditarian model. They predicted that "for those Black individuals who possess more White genes, their physical, behavioral, and other characteristics will approach those of Whites" (2005, p. 262). In the present study, we confirmed that this was the case for general cognitive ability. Moreover, we showed that the association between European ancestry and *g* was substantially mediated by eduPGS rather than skin color PGS. These results provide support for a hereditarian model.

We conducted several analyses aimed at testing a genetic hypothesis for the African-/European-American difference in cognitive ability. We found that European ancestry was a consistent predictor of cognitive ability, even after entering various controls into our models. The large observed eduPGS differences were capable of predicting substantial proportions of the cognitive ability gap, which suggests the possibility of mediation. Future assessments with cross-racially valid PGS should attempt to assess this possibility more clearly. Our global admixture-based results suggest a contribution of admixture to the African-/European-American cognitive ability gap and should encourage future investigations at the level of local admixture.

Supplementary Materials: The following are available online at http://www.mdpi.com/2624-8611/1/1/34/s1.

Author Contributions: Conceptualization, J.L., J.F., E.K.; methodology, J.L., E.K., B.P., J.F.; software, J.L. and E.K.; validation, J.L. and E.K.; formal analysis, J.L., E.K., B.P., and J.F.; investigation, J.L. and E.K.; resources, J.F. and B.P.; data curation, J.F. and B.P.; writing—original draft preparation, J.L.; writing—review and editing, J.L., E.K. and J.F.; visualization, J.L.; supervision, B.P.

Funding: This research received no external funding.

Acknowledgments: Gur, Hakonarson, and collaborators request that publications resulting from these data cite their original publication. Support for the collection of the data sets was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia. dbGaP accession phs000607.v1.p1.

Conflicts of Interest: The authors declare no conflict of interest.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

References

Adhikari, K., Chacón-Duque, J. C., Mendoza-Revilla, J., Fuentes-Guajardo, M., & Ruiz-Linares, A. (2017). The genetic diversity of the Americas. *Annual review of genomics and human genetics*, *18*, 277-296.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, *19*(9), 1655-1664.

Becker, N. S., Verdu, P., Froment, A., Le Bomin, S., Pagezy, H., Bahuchet, S., & Heyer, E. (2011). Indirect evidence for the genetic determination of short stature in African Pygmies. *American journal of physical anthropology*, *145*(3), 390-401.

Bouchard, T. J. (2013). The Wilson effect: the increase in heritability of IQ with age. *Twin Research and Human Genetics*, *16*(5), 923-930.

Bouchard Jr, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of neurobiology*, *54*(1), 4-45.

Briley, D. A., & Tucker-Drob, E. M. (2013). Explaining the increasing heritability of cognitive ability across development: A meta-analysis of longitudinal twin and adoption studies. *Psychological science*, *24*(9), 1704-1713.

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of african americans, latinos, and european Americans across the United States. *The American Journal of Human Genetics*, *96*(1), 37-53.

Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., ... & Gur, R. E. (2015). The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry*, *56*(12), 1356-1369.

Centerwall, B. S. (1978). The use of racial admixture as evidence in intelligence research: A critique. *Human Genetics*, *45*, 237-238.

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., ... & Walsh, S. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International: Genetics*, *35*, 123-135.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015.

Cheng, C. Y., Reich, D., Haiman, C. A., Tandon, A., Patterson, N., Elizabeth, S., ... & Kao, W. L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: a genetic admixture analysis in three US population cohorts. *PloS one*, *7*(3), e32840.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. *Journal of abnormal psychology*, *112*(4), 558.

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52-64.

Conley, D., & Fletcher, J. (2017). *The genome factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press.

Dixon, A. R., & Telles, E. E. (2017). Skin color and colorism: Global research, concepts, and measurement. *Annual Review of Sociology, 43*, 405-424.

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*(1), 21-50.

Dolan, C. V., & Hamaker, E. L. (2001). Investigating Black–White differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC and a critique of the method of correlated vectors. *Advances in psychology research*, *6*, 31-59.

Dolan, C. V., & Molenaar, P. C. (1994). Testing specific hypotheses concerning latent group differences in multi-group covariance structure analysis with structured means. *Multivariate Behavioral Research*, *29*(3), 203-222.

Duncan, L., Shen, H., Gelaye, B., Ressler, K., Feldman, M., Peterson, R., & Domingue, B. (2018). Analysis of polygenic score usage and performance across diverse human populations. *BioRxiv*, 398396.

Flores, C., Ma, S. F., Pino-Yanes, M., Wade, M. S., Pérez-Méndez, L., Kittles, R. A., ... & Garcia, J. G. (2012). African ancestry is associated with asthma risk in African Americans. *PloS one*, *7*(1), e26807.

Florez, J. C., Price, A. L., Campbell, D., Riba, L., Parra, M. V., Yu, F., ... & Reich, D. (2009). Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. *Diabetologia*, *52*, 1528-1536.

Freedman, M. L., Haiman, C. A., Patterson, N., McDonald, G. J., Tandon, A., Waliszewska, A., ... & Reich, D. (2006). Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proceedings of the National Academy of Sciences*, *103*(38), 14068-14073.

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence*, *51*, 79-97.

Fuerst, J. G., Lynn, R., & Kirkegaard, E. O. (2019). The Effect of Biracial Status and Color on Crystallized Intelligence in the US-Born African–European American Population. *Psych*, *1*(1), 44-54.

Giri, A., Edwards, T. L., Hartmann, K. E., Torstenson, E. S., Wellons, M., Schreiner, P. J., & Velez Edwards, D. R. (2017). African genetic ancestry interacts with body mass index to modify risk for uterine fibroids. *PLoS genetics*, *13*(7), e1006871.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., ... & Stefansson, K. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nature genetics*, *39*(5), 631-637.

Halder, I., Matthews, K. A., Buysse, D. J., Strollo, P. J., Causer, V., Reis, S. E., & Hall, M. H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep*, *38*(8), 1185-1193.

Halder, I., Shriver, M., Thomas, M., Fernandez, J. R., & Frudakis, T. (2008). A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Human mutation*, *29*(5), 648-658.

Hamer, D. H., & Sirota, L. (2000). Beware the chopsticks gene. *Molecular psychiatry*, *5*(1), 11-13.

Han, Y., Rand, K. A., Hazelett, D. J., Ingles, S. A., Kittles, R. A., Strom, S. S., ... & Haiman, C. A. (2016). Prostate cancer susceptibility in men of African ancestry at 8q24. *Journal of the National Cancer Institute*, *108*(7), djv431.

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research*, *51*(2-3), 257-258.

Hu, M., Lasker, J., Kirkegaard, E. O., & Fuerst, J. G. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych*, *1*(1), 240-261.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Jarvis, J. P., Scheinfeldt, L. B., Soi, S., Lambert, C., Omberg, L., Ferwerda, B., ... & Tishkoff, S. A. (2012). Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS genetics*, *8*(4), e1002641.

Jensen, A. R. (1969a). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39*, 1–123.

Jensen, A. R. (1969b). Reducing the heredity-environment uncertainty: A reply. *Harvard Educational Review*, *39*(3), 449-483.

Jensen, A. R. (1973). *Educability and group differences*. London: Methuen.

Jensen, A. R. (1981). Obstacles, Problems, and Pitfalls in Differential Psychology. In S. Scarr (Ed.), *Race, Social Class, and Individual Differences in IQ*. Hillsdale, NJ: Erlbaum.

Jensen, A. R. (1985). The nature of the Black–White difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, *8*(2), 193-219.

Jensen, A. R. (1998b). *The g factor*. Westport, CT: Praeger.

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., ... & Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature communications*, *5*(1), 3281.

Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J., & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome biology*, *19*, 1-14.

Kirkegaard, E. O., & Fuerst, J. (2017). Admixture in Argentina. *Mankind Quarterly*, *57*(4), 542.

Kirkegaard, E. O., Williams, R. L., Fuerst, J., & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych*, *1*(1), 1-25.

Kosoy, R., Qi, L., Nassir, R., Garcia, L., Allison, M., Shigeta, R., ... & Seldin, M. F. (2012). Relationship between hypertension and admixture in post-menopausal African American and Hispanic American women. *Journal of human hypertension*, *26*(6), 365-373.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, *50*(8), 1112-1121.

Leite, T. K., Fonseca, R. M., França, N. M. D., Parra, E. J., & Pereira, R. W. (2011). Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. *PloS one*, *6*(11), e27162.

Lima-Costa, M. F., Macinko, J., Mambrini, J. V. D. M., Peixoto, S. V., Pereira, A. C., Tarazona-Santos, E., & Ribeiro, A. L. P. (2016). Genomic African and Native American ancestry and Chagas disease: the Bambui (Brazil) Epigen cohort study of aging. *PLoS neglected tropical diseases*, *10*(5), e0004724.

Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176–193). Cambridge, England: Cambridge University Press.

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). Cross-group Comparisons of Intellectual Abilities. In *Race Differences in Intelligence* (pp. 170–171). San Francisco: Freeman.

Loehlin, J. C., Vandenberg, S. G., & Osborne, R. T. (1973). Blood group genes and Negro-White ability differences. *Behavior Genetics*, *3*, 263-270. New York, NY: Cambridge University Press.

Lynn, R. (2002). Skin color and intelligence in African Americans. *Population and Environment*, *23*, 365-375.

Lynn, R. (2015). *Race differences in intelligence: An evolutionary analysis* (2nd ed.). Arlington, VA: Washington Summit Publishers.

Marigorta, U. M., & Navarro, A. (2013). High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS genetics*, *9*(6), e1003566.

Marira, T. D., & Mitra, P. (2013). Colorism: Ubiquitous yet understudied. *Industrial and Organizational Psychology*, *6*(1), 103-107.

McDaniel, M. A., & Kepes, S. (2014). An Evaluation of Spearman's Hypothesis by Manipulating *g* Saturation. *International Journal of Selection and Assessment*, *22*(4), 333-342.

McQueen, M. B., Boardman, J. D., Domingue, B. W., Smolen, A., Tabor, J., Killeya-Jones, L., ... & Harris, K. M. (2015). The national longitudinal study of adolescent to adult health (add health) sibling pairs genome-wide data. *Behavior genetics*, *45*, 12-23.

Migliano, A. B., Romero, I. G., Metspalu, M., Leavesley, M., Pagani, L., Antao, T., ... & Kivisild, T. (2013). Evolution of the pygmy phenotype: Evidence of positive selection from genome-wide scans in African, Asian, and Melanesian pygmies. *Human biology*, *85*(1-3), 251-284.

Mollon, J., Knowles, E. E., Mathias, S. R., Gur, R., Peralta, J. M., Weiner, D. J., ... Glahn, D. C. (2018). Genetic influence on cognitive development between childhood and adulthood. *Molecular Psychiatry, 1*.

Moore, T. M., Reise, S. P., Gur, R. E., Hakonarson, H., & Gur, R. C. (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology*, *29*(2), 235.\

Murray, C. (1999, July). The Secular Increase in IQ and Longitudinal Changes in the Magnitude of the Black-White Difference: Evidence from the NLSY. In *Behavior Genetics Association Meeting*.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. New York, NY: WW Norton & Company.

Paiva, S. G. (2017). Fatores de risco para doenças cardiovasculares em quilombos contemporâneos do Brasil Central: parâmetros demográficos, socioeconômicos, ancestralidade genética e saúde. Ph.D. Thesis, University of Brasília, Brasília, Brazil.

Parra, E. J., Kittles, R. A., & Shriver, M. D. (2004). Implications of correlations between skin color and genetic ancestry for biomedical research. *Nature genetics*, *36*(Suppl 11), S54-S60.

Pearl, J. (2012). *Causality: Models, Reasoning and Inference* (2nd ed.). New York, NY: Cambridge University Press.

Pemberton, T. J., Verdu, P., Becker, N. S., Willer, C. J., Hewlett, B. S., Le Bomin, S., ... & Heyer, E. (2018). A genome scan for genes underlying adult body size differences between Central African hunter-gatherers and farmers. *Human Genetics*, *137*, 487-509.

Perry, G. H., Foll, M., Grenier, J. C., Patin, E., Nédélec, Y., Pacis, A., ... & Barreiro, L. B. (2014). Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. *Proceedings of the National Academy of Sciences*, *111*(35), E3596-E3603.

Pesta, B. J., Kirkegaard, E. O., te Nijenhuis, J., Lasker, J., & Fuerst, J. G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence*, *78*, 101408.

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral and brain sciences*, *14*(3), 373-386.

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from Gwas hits: A replication of previous findings using recent data. *Psych*, *1*(1), 55-75.

Piffer, D. (under review). The European-African difference in Cognitive-Educational polygenic scores is moderated by the trans-ethnic validity of the SNPs.

Reed, T. E. (1997). "The genetic hypothesis": It was not tested but it could have been. *American Psychologist*, 52(1), 77-78.

Rindermann, H., & Ceci, S. J. (2018). Parents' education is more important than their wealth in shaping their children's intelligence: Results of 19 samples in seven countries at different developmental levels. *Journal for the Education of the Gifted*, *41*(4), 298-326.

Rosenberg, N. A., Edge, M. D., Pritchard, J. K., & Feldman, M. W. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, medicine, and public health*, *2019*(1), 26-34.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297–330.

Rothhammer, F., & Llop, E. (1976). Amerindian descent and intellectual performance in Chilean university students. *Human Biology*, 455-464.

Rowe, D. C. (2005). Under the skin: On the impartial treatment of genetic and environmental hypotheses of racial differences. *American Psychologist*, *60*(1), 60.

Rowe, D. C., & Cleveland, H. H. (1996). Academic achievement in Blacks and Whites: Are the developmental processes similar?. *Intelligence*, *23*(3), 205-228.

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., ... & Gonzalez-José, R. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS genetics*, *10*(9), e1004572.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, public policy, and law*, *11*(2), 235.

Rushton, J. P., & Jensen, A. R. (2010). Race and IQ: A theory-based review of the research in Richard Nisbett's *Intelligence and how to get it*. *The Open Psychology Journal*, *3*(1).

Satterthwaite, T. D., Connolly, J. J., Ruparel, K., Calkins, M. E., Jackson, C., Elliott, M. A., ... & Gur, R. E. (2016). The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage*, *124*, 1115-1119.

Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., ... & Gur, R. E. (2014). Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage*, *86*, 544-553.

Scarr, S. (1981). *Race, social class, and individual differences in IQ*. Hillsdale, NJ: Erlbaum.

Scarr, S., Pakstis, A. J., Katz, S. H., & Barker, W. B. (1977). Absence of a relationship between degree of White ancestry and intellectual skills within a Black population. *Human genetics*, *39*, 69-86.

Scarr, S., & Weinberg, R. A. (1977). Intellectual similarities within families of both adopted and biological children. *Intelligence*, *1*(2), 170-191.

Sesardic, N. (2005). *Making sense of heritability*. Cambridge University Press.

Shockley, W. (1966). Possible Transfer of Metallurgical and Astronomical Approaches to the. Problem of Environment Versus Ethnic Heredity. *Science, 154:3747*, 428.

Shriner, D. (2017). Overview of admixture mapping. *Current protocols in human genetics*, *94*(1), 1-23.

Shuey, A. M. (1966). *The testing of Negro intelligence* (2nd ed.). New York: Social Science Press.

Sirin, S. R. (2005). The relationship between socioeconomic status and school outcomes [microform]: meta analytic review of research. *Rev Educ Res*, *75*(3), 417-53.

Spencer, A. V., Cox, A., & Walters, K. (2014). Comparing the efficacy of SNP filtering methods for identifying a single causal SNP in a known association region. *Annals of human genetics*, *78*(1), 50-61.

te Nijenhuis, J., van den Hoek, M., & Dragt, J. (2019). A meta-analysis of Spearman's hypothesis tested on Latin-American Hispanics, including a new way to correct for imperfectly measuring the construct of g. *Psych*, *1*(1), 101-122.

Templ, M., Kowarik, A., Alfons, A., & Prantner, B. (2019). VIM: Visualization and Imputation of Missing Values (Version 4.8.0). Available online: https://CRAN.R-project.org/package=VIM (accessed on 1 January 2019).

Templ, M., Kowarik, A., & Filzmoser, P. (2011). Iterative stepwise regression imputation using standard and robust methods. *Computational Statistics & Data Analysis*, *55*(10), 2793-2806.

Templeton, A. R. (2013). The genetic and evolutionary significance or human races. In *Race and intelligence* (pp. 31-56). Routledge.

Thomas, D. (2016). Racial IQ differences among transracial adoptees: fact or artifact?. *Journal of Intelligence*, *5*(1), 1.

Thompson, J. (2019). Piffer Kicks Against the Pricks. Available online: http://www.unz.com/jthompson/piffer-kicks-against-the-pricks/ (accessed on 2 May 2019).

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin*, *110*(3), 392.

Uricchio, L. H., Kitano, H. C., Gusev, A., & Zaitlen, N. A. (2019). An evolutionary compass for detecting signals of polygenic selection and mutational bias. *Evolution letters*, *3*(1), 69-79.

Valenzuela, C. Y., Pastene, C. S., & Perez, C. M. (1998). Intelligence and genetic markers in Chilean children. *Biological research*, *31*, 81-92.

Vinkhuyzen, A. A. E., Van Der Sluis, S., De Geus, E. J. C., Boomsma, D. I., & Posthuma, D. (2010). Genetic influences on 'environmental' factors. *Genes, Brain and Behavior*, *9*(3), 276-287.

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., ... & Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS genetics*, *2*(3), e41.

Woodley, M. A., Younuskunju, S., Balan, B., & Piffer, D. (2017). Holocene selection for variants associated with general cognitive ability: Comparing ancient and modern genomes. *Twin research and human genetics*, *20*(4), 271-280.

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. *Health policy and planning, 21*(6), 459-468.

Walsh, S., Chaitanya, L., Clarisse, L., Wirken, L., Draus-Barini, J., Kovatsi, L., ... & Kayser, M. (2014). Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Science International: Genetics*, *9*, 150-161.

Wang, X., & Teo, Y. Y. (2015). Trans-ethnic fine-mapping of rare causal variants. In *Assessing Rare Variation in Complex Traits* (pp. 253-261). Springer, New York, NY.

Warne, R. (2019). Genetic Ancestry and Cognitive Ability. *Manuscript submitted for publication*.

Weinberg, R. A., Scarr, S., & Waldman, I. D. (1992). The Minnesota Transracial Adoption Study: A follow-up of IQ test performance at adolescence. *Intelligence*, *16*(1), 117-135.

Williams, R. C., Long, J. C., Hanson, R. L., Sievers, M. L., & Knowler, W. C. (2000). Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. *The American Journal of Human Genetics*, *66*(2), 527-538.

Winkler, C. A., Nelson, G. W., & Smith, M. W. (2010). Admixture mapping comes of age. *Annual review of genomics and human genetics, 11*, 65-89.

Yeager, M., Orr, N., Hayes, R. B., Jacobs, K. B., Kraft, P., Wacholder, S., ... & Thomas, G. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nature genetics, 39*(5), 645-649.

Zanetti, D., & Weale, M. E. (2018). Transethnic differences in GWAS signals: A simulation study. *Annals of human genetics*, *82*(5), 280-286.

CHAPTER 7

MORE RESEARCH NEEDED: THERE IS A ROBUST CAUSAL VS. CONFOUNDING PROBLEM FOR INTELLIGENCE-ASSOCIATED POLYGENIC SCORES IN CONTEXT TO ADMIXED AMERICAN POPULATIONS

JOHN FUERST, EMIL O.W. KIRKEGAARD, AND DAVIDE PIFFER

Abstract

Amongst admixed American populations, polygenic scores for educational attainment and intelligence (eduPGS), genetic ancestry, and cognitive ability covary. We argue that this plausibly could be due to either confounding or to causally-relevant genetic differences between ancestral groups. It is important to determine which scenario is the case in order to better assess the validity of eduPGS. We investigate the robustness of the confounding vs. causal concern by examining, in detail, the relation between eduPGS, ancestry, and general cognitive ability in East Coast Hispanic and non-Hispanic samples. EduPGS predicted *g* among Hispanics ($B = 0.175$, $N = 506$) and all other groups (European: $B = 0.230$, $N = 4914$; European-African: $B = 0.215$, $N = 228$; African: $B = 0.126$, $N = 2179$) with controls for ancestry. Path analyses revealed that eduPGS, but not skin color, partially statistically explained the association between *g* and European ancestry among both Hispanics and the combined sample. Also, we were unable to account for eduPGS differences between the main ancestral populations using common tests for ascertainment bias and confounding related to population stratification. Overall, our results suggest that eduPGS derived from European samples can be used to predict *g* in American populations. However, owing to the uncertain cause of the ancestry-related differences in eduPGS, it is not yet clear how the effect of ancestry should be handled. We argue that causally-informative research is needed to determine the source of the relation between eduPGS, genetic ancestry, and cognitive ability.

Keywords: education, intelligence, polygenic scores, Hispanics, African-Americans

1. Introduction

General intelligence is perhaps the most powerful variable in social science, as it is not only measurable with relative ease but often strongly predicts numerous academic, economic, occupational, social, and health-related outcomes (Deary, 2010; Gottfredson, 2003). Owing to its relation to general human well-being in contemporary society, understanding the causes of individual differences in general intelligence is of significant social importance. Moreover, the search for the causes of these differences has led many to investigate the environmental and genetic determinants of general intelligence (Plomin et al., 2014). More recently, large-scale genome-wide association studies (GWAS) have been conducted to identify the genetic variants underlying the hereditary contribution to individual differences (see, e.g., Lee et al., 2018; Sniekers et al., 2017). At present, among Europeans, 4-10% of the variance in cognitive ability can be explained by intelligence and educational polygenic scores (eduPGS; Plomin & von Stumm, 2018).

GWAS have mostly been conducted on European-origin populations. Among certain groups (e.g., African Americans), European-based eduPGS have been found to display attenuated predictive accuracy with respect to cognitive ability (Guo, Lin & Harris, 2019; Lasker et al., 2019; Rabinowitz et al., 2019). There are several possible reasons for this attenuation. One explanation appeals to possible lower within-group heritability in non-European groups (e.g., Rabinowitz et al., 2019). This is a theoretically plausible account, since predictive accuracy is a function of heritability (Daetwyler, Villanueva & Woolliams, 2008). However, because the heritability of IQ is similar across racial and ethnic groups in the US (for a meta-analytic review, see Pesta et al., 2020), a more likely possibility is decay of linkage disequilibrium (LD), which results in different correlations between SNPs across different ancestry groups (Zanetti & Weale, 2018).

The significance of LD decay in attenuating transethnic predictive accuracy depends on the specific populations in question, as the impact of LD decay will be modified by ancestry-assortative mating, selection, admixture, genetic drift and other factors, which vary across populations. Although the predictive accuracy of European-derived eduPGS was found to be attenuated among African Americans, this did not seem to be the case in an admixed Brazilian sample (Horta, Hartwig & Victora, 2018). Though, if predictive validity is lower in unadmixed Africans than in Europeans, but an admixed population also has greater variance in polygenic scores than unadmixed Europeans, this greater variance in the admixed population can hide the lower predictive validity in Africans. Nor was it the case for Asian and non-Black Hispanic Americans (Guo, Lin & Harris, 2019). A similar pattern has been discovered for other, medically related PGS, with the accuracy of PGS being attenuated the most for African Americans (Duncan et al., 2019, Figure 2). This is likely because African Americans are primarily a sub-Saharan African group, and because sub-Saharan Africans are the continental lineage most genetically distant from Europeans (and other major races; Duncan et al., 2019). Generally, the validity of eduPGS has to be tested on a population by population basis and cannot be assumed to generalize.

Moreover, among admixed populations, evaluating the validity of PGS is complicated when ancestry, PGS, and traits of interest covary. In these situations, the covariance could result from a combination of ascertainment bias and confounding related to population stratification, or could be a result of genetic differences causally related to the trait (Lawson et al., 2020). Depending on the scenario, corrections for ancestry may either be accurate, overcorrect, or undercorrect. Thus, it is important to evaluate the reason for the intercorrelations.

In the case of admixed American populations, previous research has shown that ancestry, eduPGS, and cognitive ability scores are inter-correlated (Lasker et al., 2019). This is found when examining self-identified racial/ethnic (SIRE) groups separately. For example, Kirkegaard et al. (2019), Lasker et al. (2019), Warne (2020), and Guo, Lin and Harris (2019) found that cognitive ability was associated with admixture components within ethnic groups (e.g., self-identifying African Americans & Hispanics). These findings were robust to controls for SIRE, despite the fact that, as Fang et al. (2019, p. 764) noted, SIRE "acts as a surrogate to an array of social, cultural, behavioral, and environmental variables" and so "stratifying on SIRE has the potential benefits of reducing heterogeneity of these non-genetic variables and decoupling the correlation between genetic and non-genetic factors."

There are two obvious scenarios which could explain the covariance between ancestry, eduPGS, and cognitive ability. These are depicted in Figure 1. The first, (a), is the confounding scenario. Here, ancestry is associated with causally irrelevant eduPGS-related loci as a result of ascertainment bias and confounding related to population stratification (e.g., Kim et al., 2018; Martin et al., 2017); coincidentally, ancestry is also associated with cognitive ability by way of the environment. The environmental differences cause socioeconomic ones which, in turn, cause cognitive ability ones. In this scenario, eduPGS would have a spurious relation to *g* between groups, despite a causal one within groups. The second, (b), is a causal scenario. In this case, ancestry is associated with causally relevant eduPGS-related loci due to evolutionary history. These genetic differences cause cognitive ability ones which, in turn, lead to socioeconomic ones. In this case, eduPGS would be a component with constitutive explanatory relevance in the relation between ancestry and *g*.[1]

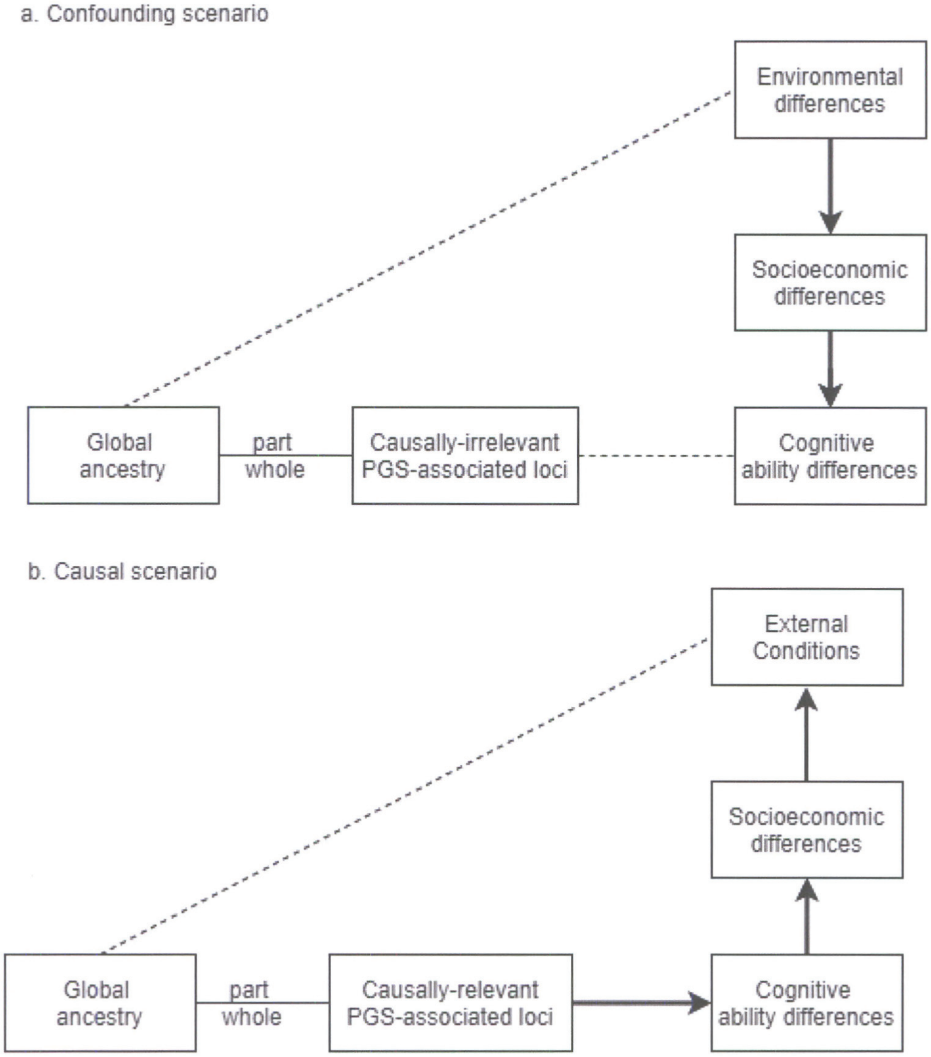

a. Confounding scenario
Environmental differences
Socioeconomic differences
Global ancestry
part
whole
Causally-irrelevant PGS-associated loci
Cognitive ability differences
b. Causal scenario
External Conditions
Socioeconomic differences
Global ancestry
part
whole
Causally-relevant PGS-associated loci
Cognitive ability differences

Figure 1. Theoretical scenarios depicting the relation between ancestry, eduPGS, cognitive ability, and socioeconomic differences

Note: In the confounding scenario (a) global or overall ancestry is correlated with environmental differences (e.g., due to vertical environmentally-transmitted factors or, perhaps, racial phenotype-based discrimination); these environmental differences cause cognitive ones by way of socioeconomic ones; at the same time, global ancestry is correlated with causally-irrelevant cognitive ability differences. In the alternative causal scenario (b), global ancestry is correlated with causally-relevant PGS-associated loci owing to evolutionary divergence in the quantitative genetic trait. These genetic differences cause cognitive ability ones which in turn cause socioeconomic ones, which lead to differences in external conditions. Note, this is not intended as a directed acyclic graph.

---

[1] Regarding terminology, since the relation between ancestry and eduPGS is better characterized as constitutive rather than causal, eduPGS could be better characterized as being a component with constitutive explanatory relevance (Craver, 2007; Ylikoski, 2013), rather than a mediator as defined by Pearl (2014).

Determining which of the two models comes closer to reality is important for an accurate interpretation of PGS in admixed American populations (Lawson et al., 2020). We do not attempt to determine which scenario is correct in this paper. Rather, we investigate if there is a robust confounding vs. causal problem with respect to cognitive ability that needs to be solved by future research. This would be the case if the data was found consistent with either a causal or confound scenario as illustrated above.

To do this, we used the Trajectories of Complex Phenotypes sample, which is a large population-representative Philadelphian sample. This sample has a number of advantages over other available ones. First, it is a local sample and so the issue of ancestry-related geographic confounding (Kirkegaard et al., 2019; Lawson et al., 2020) is not of significant concern. Second, the cognitive battery is well designed and allows for a robust examination of psychometric bias between ethnic groups. Third, the heritability of cognitive ability has been previously reported for the two largest ethnic groups (African and European Americans; Mollon et al., 2021); this is important as the predictive validity of PGS depends on the trait heritability (Pesta et al., 2020).

We focus here on the relation between ancestry, eduPGS, and general cognitive ability in the Hispanic sample and compare these results with those from European, biracial European-African, and African American samples. As far as we are aware, no previous research has investigated this issue using a largely Caribbean Hispanic sample. For background, Hispanics residing in Philadelphia are largely a Caribbean-origin, admixed population, with a predominantly Puerto Rican component (69% Puerto Rican; US Census Bureau, 2016). Hispanics, in general, have been found to score significantly below European Americans on measures of general intelligence, with Roth et al. (2017, Table 12) calculating meta-analytic effects of $d = -0.65$ to -1.04.

Given these typically large differences, we assess measurement invariance to ensure that our measure of cognitive ability functions the same for European and Hispanic Americans. Next, via regression, we examine whether associations between genetic ancestry and general cognitive ability can be accounted for by either SIRE, skin color, or parental education. Following this, we evaluate the predictive validity of eduPGS. We next examine whether the validity of eduPGS is robust to controls for genetic ancestry and color. Using path modeling, we then explore the extent to which eduPGS can statistically explain the association between European genetic ancestry and general cognitive ability.

After, we apply Jensen's Method of Correlated Vectors (MCV; Jensen, 1998) to examine the relation between eduPGS and *g*-loadings. The expectation is that eduPGS effects will be largest on the most *g*-loaded subtests since genetic effects, in contrast to environmental effects, tend to be *g*-loaded (te Nijenhuis et al., 2019). This phenomenon, of a positive correlation between the vector of *g*-loading and some other vector, is referred to as a "Jensen effect" (Woodley of Menie, Fernandes & Hopkins, 2015). While EduPGS has been found to be differentially associated with subtest scores (de la Fuente et al., 2020; Genç et al., 2021), these samples do not allow for a robust evaluation of whether eduPGS exhibits a Jensen effect owing to a limited number of subtests or small sample sizes. Thus, we evaluate the matter here. We further examine the relation between *g*-loadings, group differences, and ancestry. Since a simple causal scenario would predict a positive manifold of Jensen effects for eduPGS, heritability, ancestry, and group differences, these analyses can show if the data are consistent with such a scenario.

Finally, we examine the eduPGS scores for common forms of ascertainment bias and confounding related to population stratification to see if we can easily account for the effects of population structure. If we can, there may be no confounding vs. causal scenario in need of resolving. To be clear, though, it is outside the scope of this paper to investigate all possible forms of confounding. We therefore restrict consideration to some forms of bias discussed in the literature. The overall goal is to better understand how eduPGS, ancestry, SIRE, color, and cognitive ability are associated with one another and to evaluate whether there is a robust confound vs. causal concern in need of further research.

2. Materials and Methods

2.1. Dataset

The Trajectories of Complex Phenotypes (TCP) study was conducted by the Center for Applied Genomics at the Children's Hospital of Philadelphia, and the Brain Behavior Laboratory at the University of Pennsylvania. Participants were English-speaking Philadelphians, aged 8-21 years at the time of testing (which was done primarily between 2010 and 2013). Those with severe cognitive or medical impairments were excluded from the final sample.

2.2. Variables

The following variables were used for the admixture regression analyses:

2.2.1. Cognitive ability

Participants were administered the Penn Computerized Neurocognitive Battery (PCNB; Gur et al., 2010; Moore et al., 2015), a psychometrically robust cognitive battery that incorporates tasks linked to specific brain systems. This is a widely used 1-hour neurocognitive battery, which has been previously validated for this sample (Moore et al., 2015; Satterthwaite et al., 2016), which has an age range of 8 to 22 years. Measurement invariance for this battery was previously found to hold between African and European Americans (He & Li, 2021; Lasker et al., 2019). We created *g*-scores based on ten subtests for which measurement invariance (MI) held between our major SIRE groups (European, Hispanic, and African Americans). Details about the measures, the tests for measurement invariance, and subtest scores by subgroup can be found in Supplementary File 1.

2.2.2. Parental education

Following Lasker et al. (2019), we computed *z*-scores individually for paternal and maternal education and then averaged these. The average score was then *z*-scored again. Paternal and maternal education were the only available measures of socioeconomic status (SES). Parental education is arguably the most relevant socioeconomic related variable as it has the strongest correlation with educational-based PGS (Lee et al., 2018; Figure 4) and as it is as highly correlated with cognitive ability / scholastic achievement as are other commonly used parental-based indicators, such as occupation and income (Sirin, 2005, Table 3). That said, parental education is not a complete measure of SES, and we do not treat it as such.

2.2.3. SIRE

Self-identified race/ethnicity (SIRE) was based on yes/no questions for which the participants could select multiple races or ethnicities out of the following set of choices: Black or African-American; American Indian or Alaskan Native; Asian; European-American; Hispanic/Latino; Native Hawaiian/Pacific Islander; Other; and Not available/Pending validation. Those who identified as European only and did not report being Hispanic were coded as "European American," and mutatis mutandis for "African" and "European-African". Those who identified as Hispanic, regardless of self-identified race, were coded as such. Among Hispanics, we also identified individuals by self-identified race: European, African, Other, and a mixed category of all others (e.g., European-African Hispanic).

2.2.4. Genetic ancestry percentages

Different arrays covered different variants, so to obtain overlapping sets of single nucleotide polymorphisms (SNPs), we imputed variants with the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html). We used this server with the Minimac3 imputation algorithm, 1000G Phase 3 v5 as reference panel, and Eagle v2.3 Phasing. For computational efficiency when calculating admixture percentages, we filtered the 15.5 M variants available to the 6.5 M with a minor allele frequency (MAF) of at least 0.05. For color scores (based on a total possible of 35 variants), we did not filter by MAF, so as not to lose variants. Imputation was done using PLINK v1.90b6.8 (Chang et al., 2015). For individual ancestry estimates, we used ADMIXTURE version 1.3.0 D.H. (Alexander, Novembre & Lange, 2009). Prior, we merged the TCP with 1000 Genomes reference samples: European (British, CEU, Finnish, Spanish, and Tuscan), African (African-American SW, African-Caribbean, Esan, Gambian Mandinka, Luhya, Mende, and Yoruba), Amerindian (Peruvian), and mixed American (Puerto Rican, Colombian, and Mexican-American). We did this to gain a reliable estimate of Amerindian ancestry, since Hispanics in this sample were a three-way admixed population. We then ran ADMIXTURE with k = 3 genetic clusters. The results for the 1000 Genomes reference populations appear in Figure 2, while those for the TCP European, European-African, African, and Hispanic samples appear in Figure 3. Note that the Hispanics in the sample were largely an Afro-European descent group, with an ancestry profile matching the predominantly Puerto Rican origin of the Philadelphia Hispanic population.

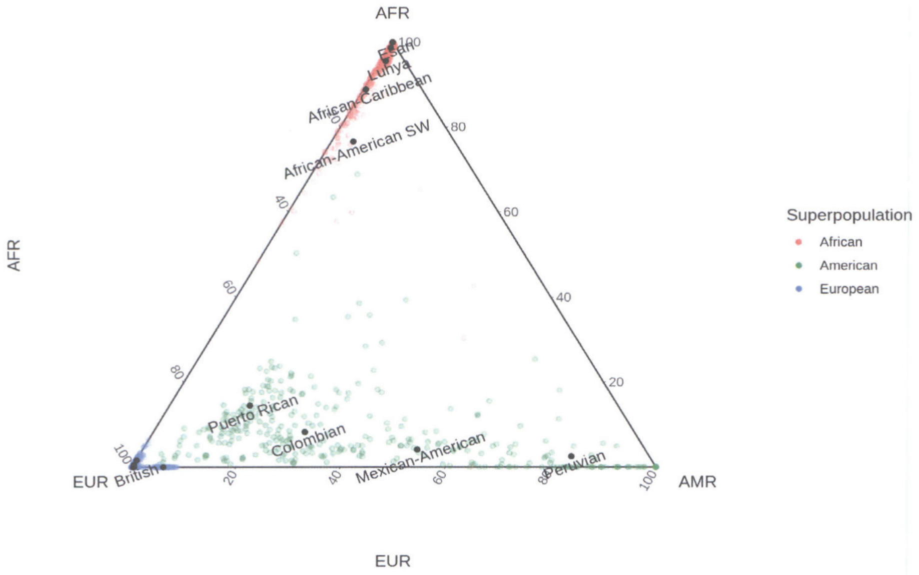


Figure 2. Admixture ternary plot for 1000 Genomes reference samples

Note: Due to clumping, some labels have been removed. Populations are: European (British, CEU, Finnish, Spanish, Tuscan), African (African-American SW, African-Caribbean, Esan, Gambian Mandinka, Luhya, Mende, Yoruba), Amerindian (Peruvian), and mixed American (Puerto Rican, Colombian, and Mexican-American).

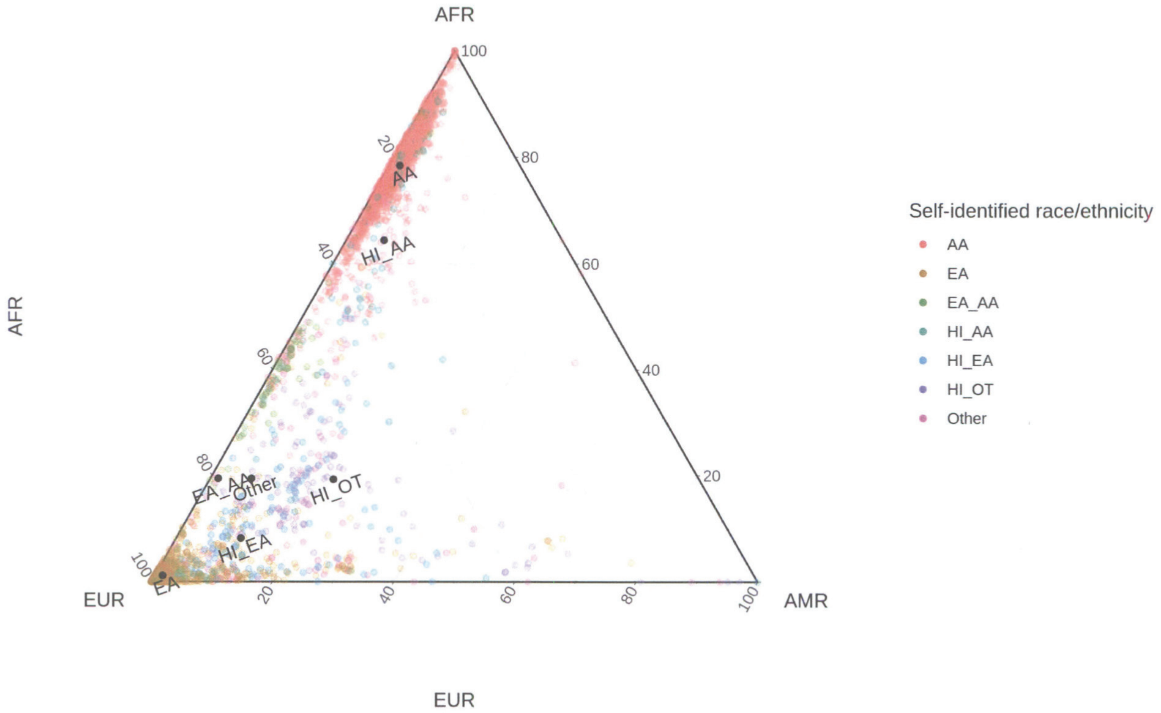


Figure 3. Admixture ternary plot for TCP self-identified race-ethnicity (SIRE) groups

Note: EA = non-Hispanic European American, EA_AA = non-Hispanic European-African American, AA = non-Hispanic African American, HI_AA= Hispanic African American, HI_EA= Hispanic European American, HI_OT = Hispanic Other, and Other = any other individual who also identified as Hispanic.

2.2.5. Skin color

Because the data did not include measures of appearance, we followed Lasker et al. (2019) and calculated phenotypic scores from genotypic data. Namely, we imputed phenotype based on genotype using the HIrisPlex-S web application (https://hirisplex.erasmusmc.nl/). Developed for use by the U.S. Department of Justice in forensic investigations, and validated on thousands of people from around the world (Chaitanya et al., 2018), the HIrisPlex-S imputes skin, hair, and eye color probabilities from 41 SNPs (with overlaps: 6 for eye color, 22 for hair color, and 36 for skin color), with a high degree of accuracy. We focus on skin color because this trait is given primacy by proponents of race-associated phenotypic discrimination ("colorism") models (e.g., Dixon & Telles, 2017; Marira & Mitra, 2013), and because the other traits have more missing data, owing to poor tagging of SNPs in some of the microarrays. A detailed discussion of the color measure and validity are provided in Supplementary File 2.

2.2.6. Cognitive ability-related polygenic scores

Lasker et al. (2019) detailed the rationale for eduPGS selection. Briefly, we validate four overlapping eduPGS from Lee et al. (2018): The eduPGS with all variants trained without the 23andMe cohort (with $N = 7{,}762{,}369$ SNPs in the present dataset), the multi-trait analysis of genome-wide association study (MTAG) eduPGS 10k SNPs (with $N = 8{,}442$ SNPs in the present dataset), the MTAG eduPGS lead SNPs (with $N = 1{,}558$ SNPs in the present dataset), and finally Lee et al.'s (2018) putatively causal variants (with $N = 111$ SNPs in the present dataset).

We used the MTAG 10k eduPGS for further analysis, since it had the highest validity in our two largest groups (European and African). This was as predicted, since both the MTAG eduPGS 10k

SNPs and MTAG–lead SNPs are predicted to show moderate within-discovery population validity and moderate validity in non-discovery populations. This is because 'lead' or 'clumped' SNPs with greater statistical significance are more likely to be causal or very close to a causal variant, which are also more likely to be transethnically valid (Grinde et al., 2019; Marigorta & Navarro, 2013; Spencer, Cox & Walters, 2014; Wang & Teo, 2015). In contrast, the eduPGS score based on all variants (without 23andMe) is predicted to show high within discovery population validity, but poor validity in non-discovery populations owing to high LD decay bias as a result of including a large number of SNPs, irrespective of their significance (Zanetti & Weale, 2018). And because there are only 111 causal SNPs, the validities are predicted to be low in both discovery and non-discovery populations (e.g., Lasker et al., 2019).

In a supplementary analysis, we additionally examined the effect of computing MTAG 10k eduPGS using population-GWAS versus within-family beta weights for the SNPs. For the population-GWAS weights, we used Lee et al.'s (2018) published MTAG weights computed based on 1.1 million individuals. For the within family weights, we contacted Lee et al. (2018), who provided us with the within family weights from their analyses of 22,000 sibling pairs. For these analyses, we computed eduPGS using the subset of MTAG SNPs for which there were both within-family and population-GWAS weights.

3. Results

3.1. Descriptive statistics

Descriptive statistics for all groups appear in Table 1. Self-identifying European, European-African, African, and Hispanic Americans were ancestrally 98%, 79%, 20%, and 60% European, respectively. Among Hispanics, those who identified as European were 81% ancestrally European, while those who identified as African were 29% ancestrally European. This difference in ancestry is consistent with those reported previously (e.g., Table S5; Bryc et al., 2015). The Hispanic Other group had the largest percentage of Amerindian ancestry, at 20%, which would be consistent with origins in Central or South America as opposed to the Caribbean (Table S5; Bryc et al., 2015).

Table 1. Sample characteristics for the participants; mean ± standard deviation, and sample size

	Age	% European	% Amerindian	Parental education	Mean *g*	Mean Color
European	13.76 ± 3.64	0.98 ± 0.07	0.01 ± 0.04	0.34 ± 0.96	0.00 ± 1.01	14.70 ± 3.84
	4937	4939	4939	4909	4914	3862
European-African	13.15 ± 3.58	0.79 ± 0.28	0.01 ± 0.03	0.16 ± 0.95	-0.14 ± 1.05	19.18 ± 7.61
	232	232	232	230	228	166
African	14.08 ± 3.75	0.20 ± 0.11	0.02 ± 0.02	-0.57 ± 0.77	-1.01 ± 1.07	30.96 ± 5.87
	2227	2228	2228	2180	217	557
Hispanic	13.59 ± 3.86	0.60 ± 0.29	0.11 ± 0.15	-0.32 ± 1.01	-0.57 ± 1.13	22.19 ± 7.59
	514	515	515	507	506	398
HI_ European	13.62 ± 4.09	0.81 ± 0.14	0.11 ± 0.09	-0.17 ± 1.03	-0.33 ± 1.17	19.45 ± 5.60
	117	117	117	115	115	80
HI_Other	13.39 ± 3.94	0.60 ± 0.16	0.20 ± 0.18	-0.53 ± 0.98	-0.65 ± 1.17	21.14 ± 6.50
	121	122	122	121	120	103
HI_African	13.88 ± 3.76	0.29 ± 0.20	0.06 ± 0.11	-0.52 ± 0.88	-0.84 ± 0.98	28.45 ± 6.78
	151	151	151	148	146	116
Other, not Hispanic	13.42 ± 3.72	0.77 ± 0.27	0.09 ± 0.16	-0.02 ± 1.06	-0.39 ± 1.13	18.17 ± 6.31
	125	125	125	123	125	99

3.2. Regression plot of the relation between g and European genetic ancestry in the combined sample

The association between cognitive ability and European ancestry for the combined sample is depicted in Figure 4. A Bayesian Generalized Additive Model line (blue) is superimposed on a regression line (orange); moreover, self-identified racial groups are color coded. As can be seen, the association is nearly linear. Additionally, European ancestry is positively and significantly associated with *g* in the admixed SIRE groups ($r_{\text{European-African}} = .26$, $r_{\text{African}} = .084$, $r_{\text{Hispanic}} = .30$), though not for European Americans ($r_{\text{European}} = .02$) among whom there is little non-European admixture.

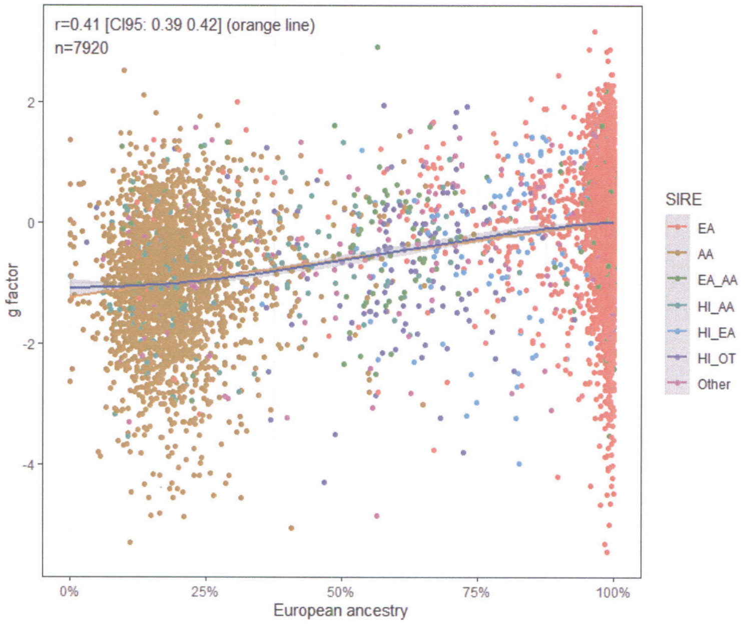


Figure 4. Regression plot of the relation between *g* and European genetic ancestry in the combined sample

Note: EA = non-Hispanic European American; HI_AA = Hispanic African American, HI_EA = Hispanic European American, HI_OT = Hispanic Other, and Other = any other individuals who also identified as Hispanic.

3.3. Regression Analysis for Ancestry as a Predictor of g

Multiple regression analysis is preferable to bivariate analysis, since there is a possibility of confounding with social and environmental factors, particularly ones correlated with SIRE (Fang et al., 2019) and race-associated phenotype (Conley & Fletcher, 2017), and also of a non-independence of admixture components. As such, we relegate the bivariate results to Supplementary File 3.

In the initial regression analysis, shown in Table 2, we deal exclusively with those who identify as Hispanic American. In the first, Model (1a), we include only African and Amerindian ancestry as independent variables. In Model 1b, we add self-identified race to see if ancestry retains predictivity and has independent explanatory power. In Model 2, we approach the issue from the perspective of "colorism" (that is, discrimination based on race-associated phenotype, specifically color). As such, in Model 2a, we include only color. We then add ancestry in 2b to see which variables retain validity. In Model 3 we further add parental education to the model with color and ancestry. Note, all values except genetic ancestry are standardized. This allows the B coefficients for ancestry to be interpreted as a change in a standard deviation of cognitive ability going from 0% to 100% ancestry. Doing so allows results from different samples with different variances in ancestry to be compared on the same metric.

Table 2. Regression analysis for ancestry as a predictor of *g* among Hispanic Americans with controls for skin color (Model 2b), and parental education (Model 2c) added; B coefficients with standard errors in parentheses

Predictor	Model 1	Model 1b	Model 2a	Model 2b	Model 3
Intercept	-0.140 (0.085)	0.285 (0.198)	-0.466 (0.056)	-0.037 (0.098)	0.116 (0.207)
Ancestry: AFR	-1.205*** (0.167)	-1.663*** (0.259)		-1.483*** (0.235)	-1.378*** (0.307)
Ancestry: AMER	-0.709* (0.337)	-0.675 (0.361)		-0.859* (0.407)	-0.624 (0.432)
SIRE HI_European		-0.408* (0.195)			-0.176 (0.205)
SIRE HI_Other		-0.480 (0.173)			-0.111 (0.181)
SIRE Other w/ HI		-0.366 (0.182)			-0.149 (0.193)
Skin Color			-0.192** (0.060)	0.136 (0.077)	0.093 (0.076)
Parental Education					0.263*** (0.051)
Adjusted R^2	0.091	0.099	0.024	0.111	0.165
N	506	506	391	391	338

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. Model 2a shows the results with color as an alternative predictor. AMR = Amerindian ancestry. AFR = African ancestry. Reference SIRE = HI_African.

As can be observed, African ancestry retained validity when either SIRE or color were added. In contrast, with the inclusion of SIRE, Amerindian ancestry became a nonsignificant predictor. However, the direction and magnitude of the effect was as previously reported (Kirkegaard et al., 2019). As seen in Model 2a, darker color was a significant predictor of lower cognitive ability, consistent with previously reported findings (e.g., Hu et al., 2019). However, the effect became nonsignificant and the sign reversed with the inclusion of ancestry. This is consistent with the view that the association between color and ability is simply indexing that between ancestry and ability (e.g., Hu et al., 2019; Lasker et al., 2019). Note, Model 1 and Model 2 above have different numbers, since there were fewer cases with color data; running the results with the same sample sizes (listwise deletion) did not lead to a difference of interpretation.

Next, we repeat the analysis with the other groups (EA, EA-AA, and AA) added. The results appear in Table 3, and are descriptively similar to those for the Hispanic-only sample. As before, Amerindian ancestry loses validity on inclusion of SIRE. This time, however, the effect size of Amerindian ancestry is close to zero. Statistically, this happens because among Philadelphian Europeans, the association between Amerindian ancestry and ability is non-significantly positive ($B = 0.500$, $S.E. = 0.391$, $N = 4914$, $p = .20$), while the association between African ancestry and ability is significantly negative ($B = -0.616$, $S.E. = 0.266$, $N = 4914$, $p < .05$). Why this is the case is not clear.

In Model 2a, color alone has validity, but it becomes nonsignificant and changes direction on inclusion of ancestry (2b). This remains the case when parental education is added in model 3.

Table 3. Regression analysis for ancestry as a predictor of *g* in the combined sample with controls for skin color (Model 2) and parental education (Model 3) added; B coefficient with standard error in parentheses

	Model 1a	Model 1b	Model 2a	Model 2b	Model 3
Intercept	0.009 (0.015)	0.010 (0.015)	-0.291*** (0.013)	0.016 (0.023)	-0.089 (0.023)
Ancestry: African	-1.274*** (0.032)	-0.988*** (0.108)		-1.262*** (0.073)	-0.785*** (0.129)
Ancestry: American	-0.532* (0.208)	0.043 (0.242)		-0.299 (0.253)	0.148 (0.281)
SIRE: African		-0.248* (0.088)			-0.107 (0.095)
SIRE: European/African		0.042 (0.072)			0.044 (0.081)
SIRE: HI_European		-0.263** (0.100)			-0.091 (0.115)
SIRE: HI_African		-0.208 (0.112)			-0.098 (0.121)
SIRE: HI_Other		-0.482*** (0.109)			-0.127 (0.115)
SIRE: Other, no Hisp.		-0.097 (0.076)			0.035 (0.082)
Skin Color			-0.390*** (0.014)	-0.003 (0.026)	0.002 (0.025)
Parental education					0.331*** (0.014)
Adjusted R^2	0.165	0.168	0.122	0.164	0.237
N	7920	7920	5991	59991	5940

Note: *$p<.05$, **$p<.01$, ***$p<.001$. Model 1b shows the results with color as an alternative predictor. Reference SIRE = European.

3.4. Analyses of the Validity of eduPGS

Again, we relegate discussion of the bivariate results for the four educational PGS to Supplementary File 3. And we used the MTAG 10k eduPGS for further analysis. While this variable did not have the highest validity among Hispanics or among European-Africans, it had the highest predictive validity among the two largest groups, European and African Americans (Supplementary File 3; Table S3-S6). Additionally, Lee et al. (2018; Table 4c) shows that MTAG eduPGS has higher predictive validity than GWAS based eduPGS for cognitive ability in two samples (the National Longitudinal Study of Adolescent to Adult Health and the Health and Retirement Study). Thus we are justified in focusing on the MTAG-based eduPGS.

The relation between this eduPGS and cognitive ability for the four major groups is shown in Figure 5. The slopes (B) and intercepts based on the regression equation were: Hispanic (*B* = 0.294, *S.E.* = 0.043; Intercept = -0.291; *N* = 506), European (*B* = 0.230, *S.E.* = 0.014; Intercept = -0.005; *N* = 4914), European-African (*B* = 0.284, 0.058; Intercept = 0.020; *N* = 228), and African American (*B* = 0.151, *S.E.* = 0.029; Intercept = -0.747; *N* = 2179). The slope for the Hispanic (*t*(5416) = 1.415, N.S.) and the European-African sample (*t*(5,138) = 0.905, N.S.) was not significantly steeper than the European one, though the power to detect a significant difference was relatively low. For the African American sample, the slope was significantly flatter (*t*(7,089) = 2.45, $p < 0.05$).

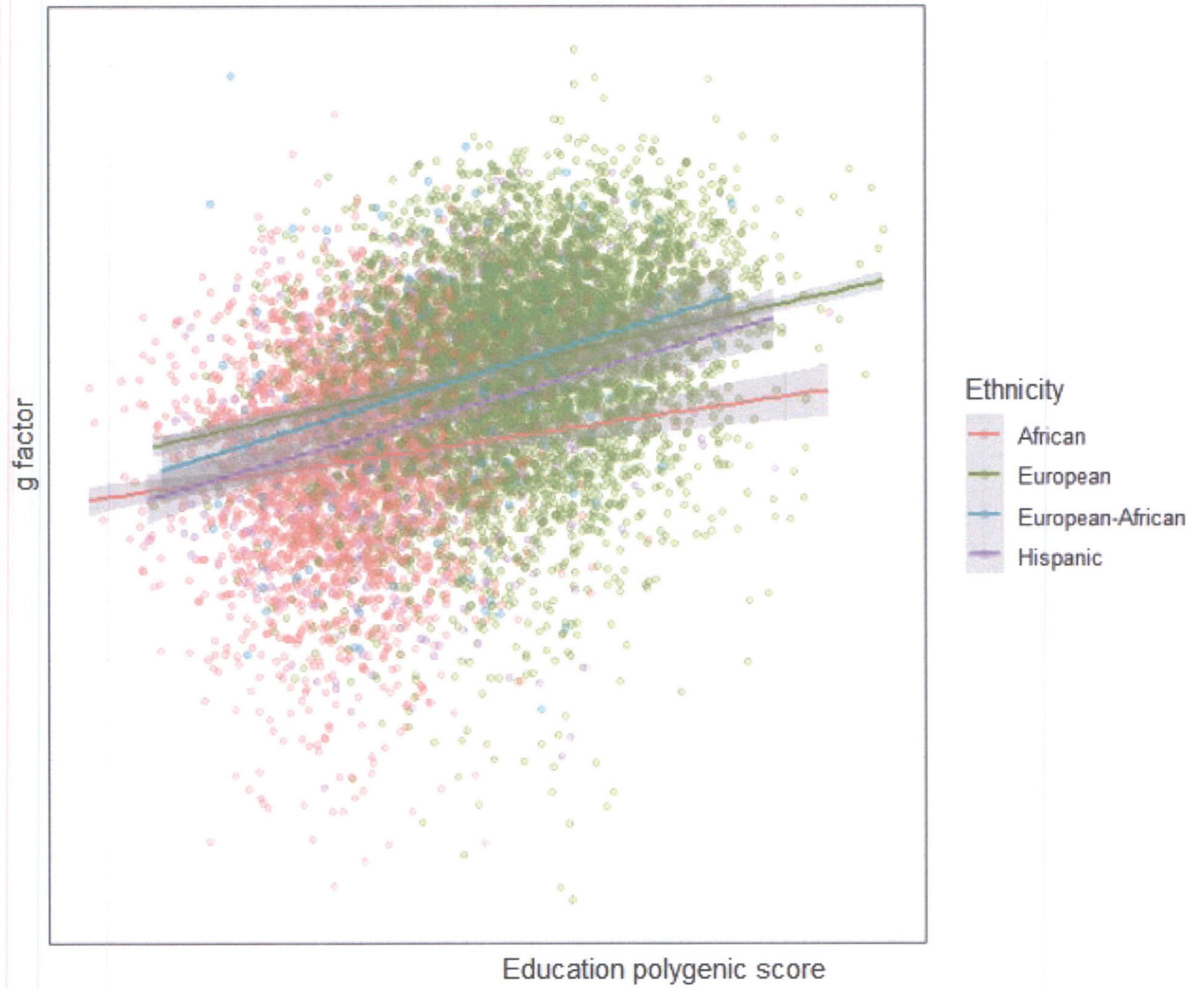


Figure 5. Regression plot for the predictive validity of MTAG 10k eduPGS with respect to *g* in the Hispanic (purple), European (green), European-African (blue), and African American (red) samples

Because we saw previously that ancestry is a robust predictor of cognitive ability in the admixed samples, we next ran multivariate regression to test to what extent the association between eduPGS and ability may be due to confounding with either global ancestry or skin color. Model 1, Table 4, shows the effect of eduPGS alone for Hispanics. EduPGS is significantly related to *g* (*B* = 0.294, *N* = 506, $p < .001$). When adding ancestry covariates in Model 2, the association attenuates but remains significant (*B* = 0.175, *N* = 506, $p < .001$). Based on the formula provided by Clogg, Petkova and Haritou (1995) for comparing betas in nested models, the *z*-score for the difference was 1.79, which is statistically significant (*p* = .037; one-tailed). Additionally, adding color in Model 3 did not further attenuate this association (*B* = 0.194, *N* = 391, $p < .001$). Note, the association in Model 2 for

Hispanics was not substantially different from that in the equivalent model for Europeans (Model 2_{European}: $B = 0.230$, $N = 4{,}914$).

Table 4. Regression results for the effect of eduPGS on cognitive ability among Hispanic Americans, B coefficient with standard error in parentheses

Predictors	**Model 1**	**Model 2**	**Model 3**
Intercept	-0.291 (0.062)	-0.115 (0.084)	-0.030 (0.096)
eduPGS	0.294*** (0.043)	0.175*** (0.051)	0.194*** (0.051)
Ancestry: African		-0.814*** (0.201)	-0.989*** (0.266)
Ancestry: Amerindian		-0.472 (0.341)	-0.492 (0.412)
Color			0.129 (0.076)
Adjusted R^2	0.084	0.110	0.141
N	506	506	391

Note: $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Next we ran the same analysis on the two other heavily admixed groups, European-African and African Americans. Table 5, Model 1 shows the effect of eduPGS alone for European-African Americans ($B = 0.284$, $N = 228$, $p < .001$). Adding ancestry in Model 2 seems to attenuate the association somewhat but the effect remains significant ($B = 0.215$, $N = 228$, $p < .01$). When color was added in Model 3, the association also remained significant ($B = 0.181$, $N = 164$, $p < .05$).

Table 5. Regression results for the effect of eduPGS on cognitive ability among European-African Americans, B coefficient with standard error in parentheses

Predictor	**Model 1**	**Model 2**	**Model 3**
Intercept	0.020 (0.074)	0.026 (0.086)	0.026 (0.124)
eduPGS	0.284*** (0.058)	0.215** (0.072)	0.181* (0.081)
Ancestry: African		-0.521 (0.296)	-0.743 (0.437)
Ancestry: Amerindian		4.247 (2.632)	5.740* (2.765)
Color			0.140 (0.130)
Adjusted R^2	0.091	0.104	0.099
N	228	228	164

Note: $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

For African Americans, shown in Table 6, the effect of eduPGS on *g* ($B = 0.151$, $N = 2{,}179$, $p < .001$) again seems to be attenuated slightly but remains significant ($B = 0.126$, $N = 2{,}179$, $p < .001$) with the inclusion of ancestry in Model 2. As seen in Model 3, adding color did not further attenuate this association ($B = 0.133$, $N = 1{,}526$, $p < .001$).

Table 6. Regression results for the effect of eduPGS on cognitive ability among African Americans, B coefficient with standard error in parentheses, B coefficient with standard error in parentheses

Predictor	**Model 1**	**Model 2**	**Model 3**
Intercept	-0.747 (0.055)	-0.423 (0.171)	-0.229 (0.204)
eduPGS	0.151*** (0.029)	0.126*** (0.030)	0.133*** (0.037)
Ancestry: African		-0.478* (0.219)	-0.532 (0.278)
Ancestry: Amerindian		0.398 (1.198)	-0.455 (1.691)
Color			-0.066 (0.043)
Adjusted R^2	0.012	0.014	0.021
N	2179	2179	1526

Note: *$p < .05$, **$p < .01$, ***$p < .001$

Finally, for the combined sample, Model 1, Table 7, shows the effect of eduPGS alone. EduPGS is significantly related to *g* ($B = 0.384$, $N = 7920$, $p < .001$). When adding ancestry covariates in Model 2, the association is attenuated but remains significant ($B = 0.221$, $N = 7{,}920$, $p < .001$). Additionally, adding SIRE and color, in Model 4, did not further substantially attenuate this association ($B = 0.218$, $N = 5{,}991$, $p < .001$).

Table 7. Regression results for the effect of eduPGS on cognitive ability for the combined sample

Predictor	Model 1	Model 2	Model 3	Model 4
Intercept	-0.322 (0.012)	-0.117 (0.016)	-0.117 (0.017)	-0.144 (0.025)
eduPGS	0.384*** (0.010)	0.221*** (0.013)	0.219*** (0.013)	0.218*** (0.014)
Ancestry: African		-0.795*** (0.042)	-0.511*** (0.110)	-0.501*** (0.136)
Ancestry: Amerindian		-0.251 (0.205)	0.214 (0.238)	0.430 (0.289)
SIRE: AA			-0.249** (0.087)	-0.224* (0.097)
SIRE: EA-AA			0.074 (0.071)	0.048 (0.084)
SIRE: Hispanic AA			-0.183 (0.110)	-0.213 (0.124)
SIRE: Hispanic EA			-0.192 (0.098)	-0.180 (0.117)
SIRE: Hispanic Other			-0.406* (0.107)	-.309** (0.117)
SIRE: Other			-0.081 (0.074)	-0.048 (0.083)
Color				0.006 (0.026)
Adjusted R^2	0.159	0.195	0.196	0.196
N	7920	7920	7920	5991

Note: $*p < .05$, $**p < .01$, $***p < .001$. SIRE (self-identified race/ethnicity) categories: AA = African American, EA = European American.

3.5. Path analysis

While fitting cross-sectional data to a path model cannot prove the causal assumptions, doing so can provide estimates of the effect magnitudes on the assumption that the model is correct (Bollen & Pearl, 2013). As such, we depict two sets of path model results fit with the **lavaan** R package (Rosseel, 2012). We limited the first path analysis to Hispanics to reduce transethnic bias in eduPGS validity. In the first model, we include European ancestry, color, and eduPGS as covariates. As color scores were not missing at random, we did not impute data, but rather handled missing data with listwise deletion. As with previous analyses, European ancestry was left unstandardized. The path model is shown in Figure 6. The detailed path estimates are shown in Supplementary File 4 (Table S1). In the model, eduPGS partially explained the association between European ancestry and cognitive ability. Independent of eduPGS, European ancestry was also strongly positively associated with cognitive ability. As expected, European ancestry was strongly negatively associated with darker color. However, darker skin color had no significant independent effect on cognitive ability. Moreover, the sign of the beta here was in the "wrong" direction relative to predictions from a colorism model. That is, darker color was unexpectedly associated with higher intelligence when controlling for ancestry. The model indicates that eduPGS is a plausible component; color, in contrast, does not seem to be a plausible mediator.

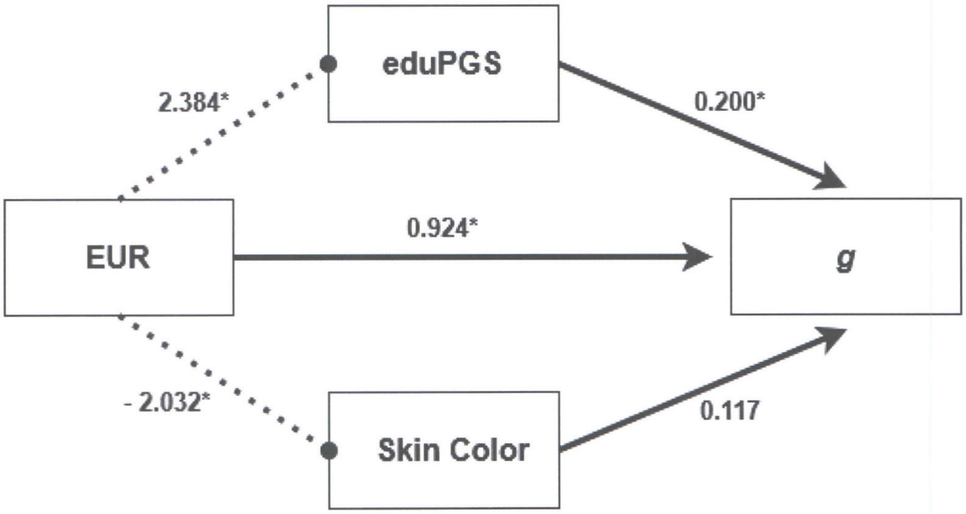


Figure 6. Path diagram for the relation between European ancestry (EUR), skin color, education polygenic score (eduPGS), and *g* in the Hispanic American sample. *N*=391. *Statistically significant at *p*<.001. Dashed-lines designate covariance

In Figure 7, we repeat this analysis with the complete sample. Since SIRE had little consistent effect, independent of ancestry, we do not include SIRE variables in the path analysis. The results are comparable, except that the beta for color is $\beta = -.001$ instead of $\beta = .117$ (both nonsignificant). Detailed results are provided in Table S2 of Supplementary File 4.

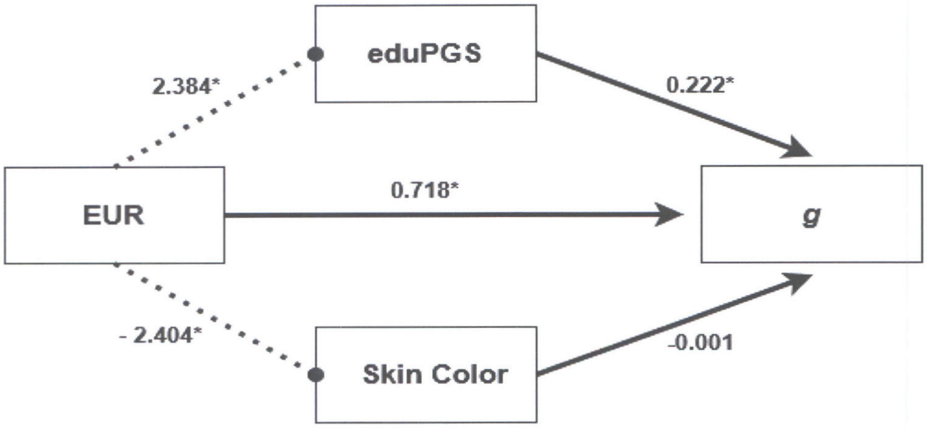


Figure 7. Path diagram for the relation between European ancestry (EUR), skin color, education polygenic score (eduPGS), and *g* in the complete sample. *N*=5,991. * Statistically significant at *p*< .001. Dashed lines designate covariance

As an alternative model, we include parental education instead of color. Color was dropped as it was not a significant predictor of *g* and because there were limited cases with color scores. While most parents are likely biological parents, not all are, and so the adolescent's ancestry is not exactly equivalent to the midparent ancestry in this case. As such, we represent the relationship between the adolescent's European ancestry and their parent's education as covariance (indicated by dashed lines). For Hispanics, the path model is shown in Figure 8, and the detailed estimates appear in Table S3 of Supplementary File 4. In the model, eduPGS again partially explained the association between European ancestry and *g*. However, parental education was also a significant predictor. The covariance between parental education and adolescent eduPGS was significant. With the current data, however, it is not possible to disentangle the causal paths between European ancestry, parental education, adolescent eduPGS, and adolescent *g*. This is because adolescent eduPGS approximates biological mid-parent eduPGS. And it is expected that, within populations, mid-parent eduPGS will be causally related to parental educational levels which, in turn, will be genetically correlated with adolescent *g* (see, e.g., Trzaskowski et al., 2014). Nonetheless, this model also indicates that eduPGS plausibly has constitutive explanatory relevance regarding the relation between European ancestry and *g*.

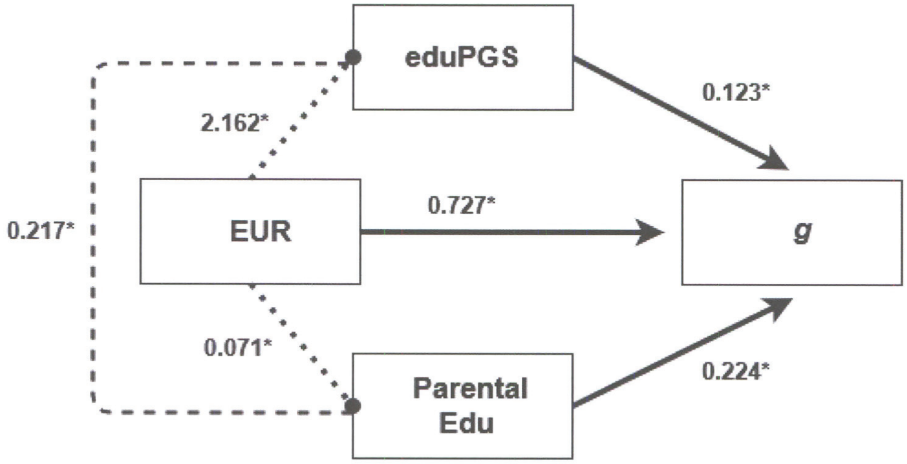


Figure 8. Path diagram for the relation between European ancestry (EUR), education polygenic score (eduPGS), parental education, and *g* in the Hispanic American sample. $N = 500$. *Statistically significant at $p < .001$. Dashed lines designate covariance

Again, we repeat this analysis with the combined sample. The path model is shown in Figure 9, and the estimates appear in Table S4 of Supplementary File 4. The results are comparable.

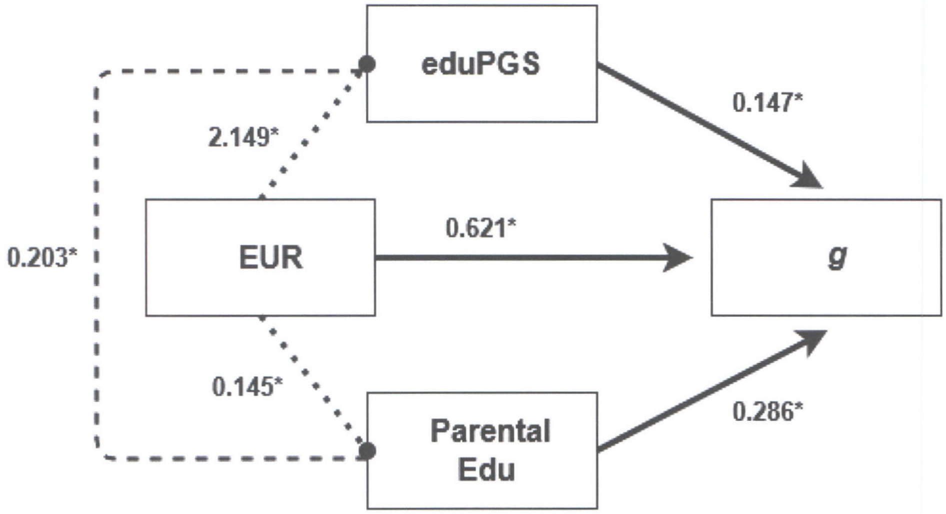


Figure 9. Path diagram for the relation between European ancestry (EUR), education polygenic score (eduPGS), parental education, and *g* in the combined sample. $N = 7846$. * Statistically significant at $p < .001$. Dashed-lines designate covariance

3.6. The Spearman-Jensen hypothesis

If the association between eduPGS and cognitive ability is primarily genetic in nature, a Jensen effect between eduPGS effects and subtest *g*-loadings is expected, because genetic effects tend to primarily act through *g* (te Nijenhuis et al., 2019). Moreover, because the weak version of Spearman's hypothesis fits the data well, for both the African-European and Hispanic-European differences, it is expected that the magnitude of the differences on subtests will positively correlate with the subtest *g*-loadings (i.e., there will be a Jensen effect on score differences). Spearman's hypothesis can also be extended to associations with genetic ancestry. If the association between ancestry (not just SIRE) and cognitive ability is primarily due to ancestry-related differences in *g*, a Jensen effect with respect to ancestry would also be expected. The subtest correlations and the subtest group differences used for this analysis are reported in Tables S5-S6 of Supplementary File 4.

Table 8 shows the vector correlations (using unrounded vectors). The results based on the ten subtests with measurement invariance appear above the diagonal; while, those for all 15 subtests appear below. As can be seen, all associations are moderately to strongly positive. The strong association between *g*-loadings and eduPGS is consistent with the finding that Lee et al.'s (2018) eduPGS is associated with genetic *g* (de la Fuente, 2020). Consistent with other research, there is a strong Jensen effect on ethnic differences (te Nijenhuis, van den Hoek & Dragt, 2019) and on ancestry-related differences within ethnic groups (Hu et al., 2019; Lasker et al., 2019). Generally, the effect of eduPGS, like that of ancestry and SIRE group differences, is pronounced on the most *g*-loaded and more heritable subtests, consistent with the predictions of a causal scenario.

Table 8. Results from Method of Correlated Vectors analysis. Correlations of subtest scores with *g* (*g* loading), with % European ancestry and with education polygenic score (PGS); size of standardized subtest gaps between SIRE groups (AA = African-American, EA = European-American, HI = Hispanic, EAA = EA-AA mixed), and subtest heritability (h^2). *N* = 10 subtests above diagonal (for the measurement invariant subtests), and 15 below diagonal (for all subtests)

	g load	Anc. *r*	Anc. (AA)	*r*Anc. (HI)	*r*PGS *r*	PGS (EA)	*r*PGS (AA)	*r*PGS (HI)	*r*EA/EAA gap	EA/AA gap	EA/HI gap	h^2
1. *g* loading	1	.89	.85	.92	.91	.90	.90	.83	.81	.94	.85	.37
2. Ancestry *r* (all)	.88	1	.69	.88	.94	.90	.73	.84	.85	.97	.99	.33
3. Ancestry *r* (AA)	.82	.62	1	.78	.71	.70	.89	.63	.87	.76	.66	.39
4. Ancestry *r* (HI)	.90	.91	.66	1	.89	.85	.85	.94	.80	.90	.82	.31
5. PGS *r* (all)	.89	.96	.65	.92	1	.99	.84	.87	.81	.96	.92	.55
6. PGS *r* (EA)	.89	.94	.65	.89	.99	1	.85	.82	.77	.93	.89	.60
7. PGS *r* (AA)	.83	.63	.84	.77	.73	.74	1	.78	.78	.82	.68	.60
8. PGS *r* (HI)	.76	.80	.53	.90	.86	.82	.72	1	.76	.87	.78	.39
9. EA/EAAA gap	.63	.66	.67	.61	.62	.59	.60	.54	1	.87	.86	.43
10. EA/AA gap	.92	.98	.69	.92	.97	.95	.72	.84	.68	1	.96	.47
11. EA/HI gap	.87	.99	.64	.86	.94	.92	.61	.75	.69	.97	1	.38
12. h^2	.43	.40	.37	.42	.56	.59	.64	.47	.30	.50	.42	1

4. Bias in Education-related PGS (eduPGS)

4.1 Evaluation of bias in the TCP sample

PGS may be biased due to the source population with which they were computed (i.e., ascertainment bias). There are a couple of obvious mechanisms by which this bias can occur. First, European-based eduPGS may be biased against non-Europeans due to the inclusion of European-specific variants (Thomson, 2019). These population-specific variants might have very low frequencies in non-European populations. Second, out-of-African based eduPGS may be biased against African populations due to an overrepresentation of derived (due to new mutations) versus ancestral (shared with other primates) variants in the out-of-Africa populations (Kim et al., 2018; Thomson, 2019). As a robustness check, we investigate both possibilities.

First, we computed MTAG eduPGS excluding variants with minor allele frequency (MAF) < .01 (leaving 7,636 overlapping variants) and < .05 (leaving 7,172 overlapping variants) among African lineages, using the 1000 Genomes reference samples to determine the African MAF. As a result, these eduPGS exclude variants not also present in African populations. As seen in Table 9, this exclusion had no substantive effect on the mean eduPGS differences between groups.

Table 9. eduPGS for European, Hispanic, and African American participants, using either the complete 10k eduPGS, or the score computed after exclusion of SNPs with minor allele frequencies of < 1% or < 5% in Africans. Standard deviations appear in parentheses

SIRE group	*N*	**eduPGS complete**	**eduPGS MAF > .01**	**eduPGS MAF > .05**
European	4939	0.02 (0.99)	0.02 (0.99)	0.02 (0.99)
European-African	232	-0.57 (1.14)	-0.55 (1.15)	-0.54 (1.14)
African	2228	-1.76 (0.80)	-1.75 (0.82)	-1.71 (0.82)
Hispanic	515	-0.94 (1.12)	-0.94 (1.13)	-0.91 (1.12)
Hisp. – European	117	-0.56 (1.08)	-0.57 (1.08)	-0.55 (1.09)
Hisp. – African	151	-1.60 (0.92)	-1.59 (0.95)	-1.54 (0.95)
Hisp. – Other	122	-0.95 (1.03)	-0.94 (1.05)	-0.90 (1.04)
Other, not Hispanic	125	-0.51 (1.08)	-0.51 (1.08)	-0.49 (1.08)

Kim et al. (2018) demonstrated that when allelic risk scores are based on out-of-Africa populations, African populations show elevated frequencies of disease-associated loci for ancestral (shared with other primates) alleles, and reduced frequencies for derived (due to new mutations after the split with primates) alleles, even when there are no underlying trait differences. They conclude that "systematic allele frequency differences between populations need not be due to any underlying difference in risk" (p. 5) and propose corrections for bias due to ancestral versus derived allele status. To investigate this, we compute eduPGS by derived and ancestral status. To be clear, we computed one PGS with only those SNPs where the enhancing allele is derived, and then another PGS with only those SNPs where the enhancing allele is ancestral. In this case, risk alleles and trait-enhancing alleles are the same thing; medical versus cognitive GWAS studies just use different terminology.

The results are shown in Table 10. As seen, contrary to the findings of Kim et al. (2018), with eduPGS, non-European populations have both lower derived and ancestral eduPGS scores. Moreover, the differences are largest for the derived ones. As a result, when Kim et al.'s (2018, p. 12) correction is applied, the polygenic score gaps change little.

Table 10. eduPGS for European, European-African, African, and Hispanic American participants, mean ± standard deviation

SIRE group	*N*	**eduPGS derived**	**eduPGS ancestral**	**eduPGS corrected***
European-American	4939	0.02 ± 0.99	0.01 ± 1.00	0.02
European-African	232	-0.68 ± 1.15	-0.34 ± 1.12	-0.51
African-American	2228	-2.05 ± 0.85	-1.14 ± 0.79	-1.60
Hispanic	515	-1.05 ± 1.21	-0.66 ± 1.02	-0.85
Hisp.: European	117	-0.57 ± 1.07	-0.45 ± 1.05	-0.49
Hisp.: Other	122	-1.01 ± 1.03	-0.72 ± 1.08	-0.85
Hisp.: African	151	-1.84 ± 1.06	-1.06 ± 0.82	-1.45
Other, not Hispanic	125	-0.60 ± 1.20	-0.31 ± 0.97	-0.46

* Corrected following Kim et al.'s (2018) procedure, which involves adjusting the ancestral PGS down by 0.1902% and adjusting the derived PGS up by 0.1082% and then averaging the two. Kim et al. report that, in general, "ancestral risk [i.e., trait enhancing] alleles are found at 9.51% higher frequency in Africa, and derived risk [i.e., trait enhancing] alleles are found at 5.40% lower frequency in Africa" (p. 1). They then applied this general finding, with percentages doubled to account for diploidy, to specific traits. We do the same.

Additionally, we computed the validities to see if the corrections affected these. The validities by eduPGS are reported in Table 11. As seen, the validities for the different eduPGS were approximately the same for a given SIRE group.

Table 11. Predictive validities of MAF ≥ 0.01, MAF ≥ 0.05, derived, and ancestral eduPGS among European, European-African, Hispanic, African and Hispanic Americans (*N* = sample size)

SIRE group	*N*	**MTAG eduPGS**	**10keduPGS MAF ≥ 0.01**	**eduPGS MAF ≥ 0.05**	**eduPGS derived**	**eduPGS ancestral**
European	4914	.227	.227	.224	.207	.189
European-African	228	.308	.305	.298	.283	.263
African	2228	.112	.113	.115	.103	.090
Hispanic	515	.293	.288	.287	.292	.225

Finally, it has been argued that the eduPGS differences may be upwardly biased by using the discovery sample-based direction of effects for SNPs (Thompson, 2019). To investigate this we computed the betas for the MTAG SNPs, used as predictors of *g*, for European (*N* = 4,939) and African Americans (*N* = 2,228) separately. The SIRE specific betas are provided in the Supplementary Material. We then examined mean differences and validities for the variants which showed transracially concordant effects across ethnic groups as compared to those which showed discordant effects. African Americans had lower eduPGS than EA based on the concordant SNPs (-

3.13 and 0.00, respectively), but slightly higher eduPGS based on the discordant SNPs (0.49 and 0.00, respectively). Moreover, among African Americans, the concordant SNPs showed higher validity than the discordant SNPs for *g*. Cross-validation confirmed that the concordant eduPGS were more predictive than the discordant eduPGS among African Americans. A possible interpretation is that the discordant eduPGS, which are less likely to be causal, contain more LD decay related effects. Regardless, there is no evidence, based on this sample, that the eduPGS differences are being inflated by the inclusion of SNPs with transethnically discordant effects as some have argued (Thompson, 2019).

4.2. Evaluation of bias in the 1000 Genomes samples

We additionally ran supplementary analyses which leveraged the 1000 Genomes data to explore the effects of score construction on the population differences. Methods and detailed results are reported in Supplementary File 5. First, we examined the effect of using population-GWAS vs. within-family PGS. In this analysis, we found that the African-European differences based on the within-family PGS were reduced in size but nonetheless large. Next, we examine the effect of using trans-ethnically concordant betas. The results were similar to those reported in our sample. Finally, we examined the predictive validity of different polygenic score constructions across populations. While the expected associations remained positive, we found that the construction of eduPGS affected the magnitude of correlations, consistent with reports by others (e.g., Berg et al., 2019, Figure 1; Sohail et al., 2019, Figure 4). These results highlight Duncan et al.'s (2019) caution that different eduPGS can give markedly different results, rendering interpretation uncertain.

5. Discussion

5.1. Conclusion

To better understand how eduPGS function in admixed American populations, and to determine if there is a robust confound vs. causal problem, we examined the association between intelligence-related polygenic scores, global ancestry, and general cognitive ability in Hispanic and non-Hispanic European, European-African, and African American samples. In this sample, we were able to confirm full factorial invariance for the cognitive tests in the case of the European-Hispanic and European-African differences. And, previously, full factorial invariance had been found to hold along the full range of European ancestry (Lasker et al., 2019). Thus, the associations found reflect those with latent mental ability.

Among Hispanics and in the combined Hispanic and non-Hispanic American sample, the association between European/African ancestry and cognitive ability was robust to controls for SIRE, color, and parental education. For Amerindian ancestry, in both the Hispanic-only and the combined samples, the association with cognitive ability became nonsignificant when SIRE was added as a covariate. In the Hispanic-only sample, the effect remained directionally consistent with previously reported results (Kirkegaard et al., 2019; Warne, 2020). Statistically, the insignificance, in this case, was a result of the higher standard errors of Amerindian ancestry, as compared to African, which was due to the lower variance in Amerindian ancestry. Similarly, in a small, "non-Hispanic Caucasian" sample, with little variance in ancestry, European ancestry versus (apparently) Mexican/Mexican-American ancestry was only weakly and non-significantly associated with IQ ($r = .09$, $N = 120$; Wang et al., 2016). For the combined sample, however, the effect of Amerindian ancestry turned positive with SIRE controls. Statistically this was due to Amerindian ancestry being slightly positively correlated with general intelligence in the non-Hispanic White sample ($r = .014$; $N = 4914$, *N.S.*) and to the much larger non-Hispanic White than Hispanic sample size. Generally, the relation between cognitive ability and Amerindian ancestry needs to be better explored in predominantly European-Amerindian "Mestizo" samples. The results to date are ambiguous.

Here, explanations for the association between European ancestry and cognitive ability by confounding with either geography or racial phenotype (see, e.g., Conley & Fletcher, 2017) are not viable. This was a local population sample from the Philadelphia area, so geographic confounding is not a substantial concern. We did find a significant negative association between cognitive ability and darker skin color. However, rather than the association between ancestry and cognitive ability being explained by color, the association between color and cognitive ability was statistically explained by ancestry.

These latter results are key in that there is a large literature on "colorism" which purports to demonstrate color or pigment-based discrimination by showing mere correlations between color and social outcomes (e.g., Marira & Mitra, 2013). Moreover, it has been argued by some that such discrimination might account for potential associations between ancestry and cognitive ability (e.g., Conley & Fletcher, 2017). However, our results concur with the competing distributional model described by Hu et al. (2019, Figure 1), in which the association between color and cognitive ability is a proxy (versus a cause) of that between ancestry and cognitive ability. This finding has implications for genetic research (Lawson et al., 2020), as it suggests that color-based discrimination is not likely an additional source of confounding.

We found that eduPGS was significantly associated with cognitive ability within the European, European-African, African, and Hispanic samples. For the MTAG 10k eduPGS, which we use for further analyses, the associations in the admixed populations were attenuated with the inclusion of ancestry. Regardless, the association remained significant and not substantially different from that in the European sample ($B_{\text{European}} = .230$ vs. $B_{\text{European-African}} = .215$ and $B_{\text{Hispanic}} = .175$), except in the case of African Americans ($B_{\text{African}} = .126$) where, while statistically significant, the association was less than that for European Americans.

Moreover, path analysis indicated that eduPGS partially statistically explained the association between European ancestry and cognitive ability, whereas color did not. Additionally, the explanatory power of eduPGS was not fully accounted for by parental education, a variable which is both genetically and environmentally correlated with adolescent intelligence within groups and thus may be likewise between groups.

The eduPGS differences may be spurious owing to difficult to control ascertainment bias and confounding related to population stratification. For example, Berg et al. (2019) found attenuated effects for height PGS when applying UK versus pan-European-based PGS to Eurasian samples. The reason seems to be that controls for ancestry components do not always fully capture population structure effects, however, some confounding effects can be avoided by computing PGS using a more homogeneous population and then applying these to the more heterogeneous populations of interest (see: Berg & Coop, 2014). In our case, though, we start with eduPGS based on European origin samples and then apply these to samples of different continental ancestry. As such, we already incorporated this component of Berg et al.'s (2019) analysis into ours.

We further investigated whether differences might be due to biasing effects of discovery population-specific variants, whether the allele associated with higher values of the trait is derived or ancestral, failure of SNP sign concordance between populations, and discrepancy between population-GWAS and within-family coefficients. The forms of ascertainment bias and confounding related to population stratification addressed by our procedures do not explain the group differences in eduPGS. However, it is beyond the scope of this paper to explore all forms of possible confounding.

The regression and path analysis results raise some possible concerns regarding the use and interpretation of eduPGS in admixed American populations. Ancestry covaries with trait and eduPGS scores. This could either be due to trait-relevant genetic differences between the ancestral groups of the admixed populations or to a mix of both environmental differences and also ascertainment bias and confounding related to population stratification. If the former, controlling for ancestry can attenuate the effect of eduPGS; however, if the latter, leaving ancestry unadjusted can inflate it.

Additionally, MCV indicated that the effect of eduPGS was strongly *g*-loaded, as was the effect of ancestry and group differences. This is consistent with all these effects acting primarily by way of *g*. These results for eduPGS are not self-evident. This is because it has been shown that some of the effect of eduPGS is a shared environmental effect (Domingue & Fletcher, 2019); however, the effect of adoption, a shared environmental effect, exhibits an anti-Jensen effect (te Nijenhuis, Jongeneel-Grimen & Armstrong, 2015). Thus, that eduPGS would act like typical genetic effects, and be most pronounced on *g*-loaded subtests, is not obvious. In addition to eduPGS, both ancestry and group differences exhibited Jensen effects, which is consistent with differences being primarily in *g* (i.e., Spearman's hypothesis). A practical implication of this is that good measures of *g* are needed to capture the full statistical effects of cognitive ability in context to research on eduPGS and related variables. Moreover, the positive manifold of Jensen effects is at least consistent with a causal model. To reconcile it with a confounding model one would need to additionally propose a mechanism by which environmentally induced phenotypic differences produced Jensen effects (e.g., Flynn, 2019).

The differences in eduPGS between ancestral groups are large enough that this is an issue which future research should try to resolve. While the magnitudes of differences for the true causal SNPs are unknown, magnitudes can be calculated for presently known education-related SNPs. We can use the fixation index, a measure of population differentiation, to do this. Supplementary File 6 shows the Fst for the SNPs from Lee et al. (2018) for 1000 Genomes super-populations. The Fst value between Europeans and Africans, which are the two main ancestral groups for the admixed populations here, is .1090 As detailed in Supplementary File 3, for typically reported heritabilities ($h^2 = .5$; Pesta et al., 2020; Polderman et al., 2015) this magnitude of population differentiation gives medium to large expected mean differences (i.e., when environments are equal), in this case, equivalent to $d = 0.68$. Thus, if transethnically unbiased eduPGS differences turn out to be commensurate with those based on Lee et al.'s (2018) eduPGS, medium to large phenotypic differences are expected under conditions of environmental equality. Controlling for ancestry might then bias eduPGS effects.

Related to this point, a reader suggested that we should run analyses to detect polygenic selection as done by Berg et al. (2019). However, whether differences are due to drift or selection is not necessarily relevant to whether there are genetic differences between populations. Moreover, given the expected differences discussed above, selection, in the form of stabilizing or convergent selection *between populations* is needed to show trait equality, not trait differences. That is, the evolutionary default or null expectation would be that trait-causing SNP frequency differences will be commensurate with random SNP ones, not that selection acted to homogenize differences in this particular trait (Edelaar & Björklund, 2011; Edge & Rosenberg, 2015; Leinonen et al., 2013; Rosenberg et al., 2019). As Rosenberg et al. note: "[P]henotypic differences among populations are predicted under neutrality to be similar in magnitude to typical genetic differences among populations" (*p*. 30). We are not aware of anyone who has established selection acting between populations such as to make causally-relevant genetic differences (related to education and intelligence) substantially smaller than expected by drift. Whether there was divergent selection is undetermined (e.g., Guo et al., 2018; Racimo, Berg & Pickrell, 2018).

Considering these results, it may be that individual ancestry tracks cognitive ability in admixed populations across the Americas. If so, proper interpretation of the predictive accuracy of eduPGS in American admixed populations will require a better understanding of the causal pathways underwriting this association. We argue that admixture mapping is the appropriate next step (for a rationale, see Kirkegaard et al., 2019). That said, different sub-ancestral components (e.g., North vs. South Amerindian / European, West versus East African) may yield different associations between global ancestry and cognitive ability, so additional global ancestry studies are warranted to better understand the pattern of effects.

5.2. Limitations

Owing to the sample sizes of the subgroups, some of the ancillary analyses, such as the comparison of eduPGS validities, were underpowered. Replications should be attempted with larger samples; until then, caution is warranted regarding interpretation. Importantly, we did not attempt to determine the cause of the covariance between eduPGS, ancestry, and *g*. We can only say that the cause is presently undetermined and that it needs to be resolved for a proper interpretation of the predictive accuracy of eduPGS in admixed American populations. While we suggest local admixture mapping as a way to narrow the uncertainty, further research should also explore other forms of confounding.

References

Alexander, D.H., Novembre, J. & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655-1664.

Berg, J.J. & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics* 10(8): e1004412.

Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., ... & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* 8: e39725.

Bollen, K.A. & Pearl, J. (2013). Eight myths about causality and structural equation models. In: S.L. Morgan (ed.), *Handbook of Causal Analysis for Social Research*, pp. 301-328. Dordrecht: Springer.

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics* 96: 37-53.

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., ... & Kayser, M. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International: Genetics* 35: 123-135.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4(1): 7.

Clogg, C.C., Petkova, E. & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100: 1261-1293.

Conley, D. & Fletcher, J. (2017). *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton, NJ: Princeton Univ. Press.

Craver, C. (2007). Constitutive explanatory relevance. *Journal of Philosophical Research* 32: 3-20.

Daetwyler, H.D., Villanueva, B. & Woolliams, J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3(10): e3395.

de la Fuente, J., Davies, G., Grotzinger, A.D., Tucker-Drob, E.M. & Deary, I.J. (2020). A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nature Human Behaviour* 5: 49-58.

Deary, I.J. (2010). Cognitive epidemiology: Its rise, its current issues, and its challenges. *Personality and Individual Differences* 49: 337-343.

Dixon, A.R. & Telles, E.E. (2017). Skin color and colorism: Global research, concepts, and measurement. *Annual Review of Sociology* 43: 405-424.

Domingue, B.W. & Fletcher, J. (2019). Separating measured genetic and environmental effects: Evidence linking parental genotype and adopted child outcomes. *Behavior Genetics* 50: 301-309.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., ... & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications* 10(1): 3328.

Edelaar, P.I.M. & Björklund, M. (2011). If F_{ST} does not measure neutral genetic differentiation, then comparing it with QST is misleading. Or is it? *Molecular Ecology* 20: 1805-1812.

Edge, M.D. & Rosenberg, N.A. (2015). Implications of the apportionment of human genetic diversity for the apportionment of human phenotypic diversity. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 52: 32-45.

Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., ... & DuVall, S.L. (2019). Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *American Journal of Human Genetics* 105: 763-772.

Flynn, J.R. (2019). A response to te Nijenhuis et al. (2019). *Journal of Biosocial Science* 51: 913-916.

Genç, E., Schlüter, C., Fraenz, C., Arning, L., Metzen, D., Nguyen, H.P., ... & Ocklenburg, S. (2021). Polygenic scores for cognitive abilities and their association with different aspects of general intelligence – a deep phenotyping approach. *Molecular Neurobiology*: https://link.springer.com/article/10.1007/s12035-021-02398-7.

Gottfredson, L.S. (2003). *g*, jobs, and life. In: H. Nyborg (ed.), *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen*, pp. 293-342. New York: Pergamon.

Grinde, K.E., Qi, Q., Thornton, T.A., Liu, S., Shadyab, A.H., Chan, K.H.K., ... & Sofer, T. (2019). Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genetic Epidemiology* 43: 50-62.

Guo, G., Lin, M.J. & Harris, K.M. (2019). Socioeconomic and genomic roots of verbal ability. *bioRxiv*. https://www.biorxiv.org/content/biorxiv/early/2019/02/10/544411.full.pdf

Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., ... & Yang, J. (2018). Global genetic differentiation of complex traits shaped by natural selection in humans. *Nature Communications* 9: 1865.

Gur, R.C., Richard, J., Hughett, P., Calkins, M.E., Macy, L., Bilker, W.B., ... & Gur, R.E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods* 187: 254-262.

He, Q. & Li, J.J. (2021). Factorial invariance in hierarchical factor models of mental disorders in African American and European American youths. *Journal of Child Psychology and Psychiatry* 62: 289-298.

Horta, B.L., Hartwig, F.P. & Victora, C.G. (2018). Breastfeeding and intelligence in adulthood: Due to genetic confounding? *Lancet Global Health* 6(12): e1276-e1277.

Hu, M., Lasker, J., Kirkegaard, E.O. & Fuerst, J.G. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych* 1(1): 240-261.

Jensen, A.R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.

Kim, M.S., Patel, K.P., Teng, A.K., Berens, A.J. & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome Biology* 19: 179.

Kirkegaard, E.O., Woodley of Menie, M.A., Williams, R.L., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych* 1(1): 1-25.

Lasker, J., Pesta, B.J., Fuerst, J.G. & Kirkegaard, E.O. (2019). Global ancestry and cognitive ability. *Psych* 1(1): 431-459.

Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., ... & Timpson, N.J. (2020). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* 139: 23-41.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Fontana, M.A. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics* 50: 1112-1121.

Leinonen, T., McCairns, R.S., O'Hara, R.B. & Merilä, J. (2013). Q_{ST}–F_{ST} comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics* 14: 179-190.

Marigorta, U.M. & Navarro, A. (2013). High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genetics* 9(6): e1003566.

Marira, T.D. & Mitra, P. (2013). Colorism: Ubiquitous yet understudied. *Industrial and Organizational Psychology* 6(1): 103-107.

Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., ... & Kenny, E.E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *American Journal of Human Genetics* 100: 635-649.

Mollon, J., Knowles, E.E., Mathias, S.R., Gur, R., Peralta, J.M., Weiner, D.J., ... & Glahn, D.C. (2021). Genetic influence on cognitive development between childhood and adulthood. *Molecular Psychiatry* 26: 656-665.

Moore, T.M., Reise, S.P., Gur, R.E., Hakonarson, H. & Gur, R.C. (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology* 29: 235-246.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods* 19: 459-481.

Pesta, B.J., Kirkegaard, E.O., te Nijenhuis, J., Lasker, J. & Fuerst, J.G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence* 78: 101408.

Plomin, R. & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics* 19: 148-159.

Plomin, R., DeFries, J.C., Knopik, V.S. & Neiderhiser, J.M. (2014). *Behavioral Genetics*, 6th ed. New York: Worth Publishers.

Polderman, T.J., Benyamin, B., De Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M. & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47: 702-709.

Rabinowitz, J.A., Kuo, S.I.C., Felder, W., Musci, R.J., Bettencourt, A., Benke, K., ... & Kouzis, A. (2019). Associations between an educational attainment polygenic score with educational attainment in an African American sample. *Genes, Brain and Behavior* 18(5): e12558.

Racimo, F., Berg, J.J. & Pickrell, J.K. (2018). Detecting polygenic adaptation in admixture graphs. *Genetics* 208: 1565-1584.

Rosenberg, N.A., Edge, M.D., Pritchard, J.K. & Feldman, M.W. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and Public Health* 2019: 26-34.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(2): 1-36. http://www.jstatsoft.org/v48/i02/

Roth, P.L., van Iddekinge, C.H., DeOrtentiis, P.S., Hackney, K.J., Zhang, L. & Buster, M.A. (2017). Hispanic and Asian performance on selection procedures: A narrative and meta-analytic review of 12 common predictors. *Journal of Applied Psychology* 102: 1178-1202.

Satterthwaite, T.D., Connolly, J.J., Ruparel, K., Calkins, M.E., Jackson, C., Elliott, M.A., ... & Qiu, H. (2016). The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage* 124: 1115-1119.

Sirin, S.R. (2005). The relationship between socioeconomic status and school outcomes: Meta-analytic review of research. *Review of Educational Research* 75: 417-453.

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R., Krapohl, E., ... & Amin, N. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics* 49: 1107-1112.

Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., ... & Sunyaev, S.R. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife 8: e39702.

Spencer, A.V., Cox, A. & Walters, K. (2014). Comparing the efficacy of SNP filtering methods for identifying a single causal SNP in a known association region. *Annals of Human Genetics* 78: 50-61.

te Nijenhuis, J., Jongeneel-Grimen, B. & Armstrong, E.L. (2015). Are adoption gains on the g factor? A meta-analysis. *Personality and Individual Differences* 73: 56-60.

te Nijenhuis, J., van den Hoek, M. & Dragt, J. (2019). A meta-analysis of Spearman's Hypothesis tested on Latin-American Hispanics, including a new way to correct for imperfectly measuring the construct of *g*. *Psych* 1(1): 101-122.

te Nijenhuis, J., Choi, Y.Y., van den Hoek, M., Valueva, E. & Lee, K.H. (2019). Spearman's hypothesis tested comparing Korean young adults with various other groups of young adults on the items of the Advanced Progressive Matrices. *Journal of Biosocial Science* 51: 875-912.

Thompson, J. (2019, May 2). Piffer kicks against the pricks. Available online: http://www.unz.com/jthompson/piffer-kicks-against-the-pricks/

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., ... & Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence* 42: 83-88.

US Census Bureau (2016). *American factfinder – Results*. US Department of Commerce, Economics and Statistics Administration. Retrieved from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

Wang, F.L., Pandika, D., Chassin, L., Lee, M. & King, K. (2016). Testing the relations among family disorganization, delay discounting, and adolescent alcohol use: A genetically informed study. *Alcoholism: Clinical and Experimental Research* 40: 846-856.

Wang, X. & Teo, Y.Y. (2015). Trans-ethnic fine-mapping of rare causal variants. In: E. Zeggini & A. Morris (eds.), *Assessing Rare Variation in Complex Traits*, pp. 253-261. New York: Springer.

Warne, R.T. (2020). Continental genetic ancestry source correlates with global cognitive ability score. *Mankind Quarterly* 60: 400-422.

Woodley of Menie, M.A.W., Fernandes, H.B. & Hopkins, W.D. (2015). The more *g*-loaded, the more heritable, evolvable, and phenotypically variable: Homology with humans in chimpanzee cognitive abilities. *Intelligence* 50: 159-163.

Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis* 78(2): 277-297.

Zanetti, D. & Weale, M.E. (2018). Transethnic differences in GWAS signals: A simulation study. *Annals of Human Genetics* 82(5): 280-286.

CHAPTER 8

LINEAR AND PARTIALLY LINEAR MODELS OF BEHAVIORAL TRAIT VARIATION USING ADMIXTURE REGRESSION

GREGORY CONNOR AND JOHN G. R. FUERST

Abstract

The admixture regression methodology exploits the natural experiment of random mating between individuals with different ancestral backgrounds to infer the environmental and genetic components to trait variation across racial and ethnic groups. This paper provides a statistical framework for admixture regression based on the linear polygenic index model and applies it to neuropsychological performance data from the Adolescent Brain Cognitive Development (ABCD) database. We develop and apply a new test of the differential impact of multi-racial identities on trait variation, an orthogonalization procedure for added explanatory variables, and a partially linear semiparametric functional form. We find a statistically significant genetic component to neuropsychological performance differences across racial identities, and find some possible evidence of nonlinearity in the link between admixture and neuropsychological performance scores in the ABCD data.

Keywords: admixture regression, semiparametric model, linear polygenic index

1. Introduction

Racial/ethnic group identities such as Black, White, Hispanic, Native American, East Asian and South Asian show empirically strong linkages to medical and behavioral traits such as obesity (Wang & Beydoun, 2007), type 2 diabetes (Cheng et al., 2013), hypertension (Lackland, 2014), asthma (Choudhry et al., 2006), neuropsychological performance (Llibre-Guerra et al., 2018), smoking behaviors (Choquet, Yin, & Jorgenson, 2021), and sleep disorders (Halder et al., 2015). An important research question is to what degree any such observed trait variation arises from differences in the typical diets, cultural practices and other environmental particularities of the racial/ethnic groups, or from similarity in genetic pools within each group traceable to shared geographic ancestry. Many diverse national populations descend demographically from isolated continental groups within a few hundred years. Modern genetic technology can measure with high accuracy the proportion of an individual's ancestry associated with these continental groups. Also, in many culturally diverse nations, most individuals can reliably self-identify as members of one or more racial or ethnic groups. Admixture regression leverages these two data sources, self-identified race or ethnicity (SIRE) and genetically-measured admixture proportions, to decompose trait variation correspondingly. Admixture regression has been widely applied to medical and behavioral traits including asthma (Salari et al., 2005), body mass index (Klimentidis, Miller, & Shriver, 2009), type 2 diabetes (Cheng et al., 2013), blood pressure (Klimentidis et al., 2012), neuropsychological performance (Llibre-Guerra et al., 2018), and sleep depth (Halder et al., 2015). It has particular value in the case of complex behavioral traits where reliably identifying genetic loci associated with trait variation is beyond the current reach of science. Admixture mapping is a more technically challenging methodology, often used in conjunction with admixture regression, which uses ancestral population trait differences to attempt to identify genetic loci associated with a trait. This paper focusses exclusively on admixture regression.

This paper first develops a simple statistical framework for admixture regression of behavioral traits by linking it to the linear polygenic index model from behavioral genetics; this framework clarifies the key assumptions that are implicit in this simple and powerful statistical technique. The paper then extends the admixture regression methodology in several ways. We provide a new test statistic for identifying whether a given multi-racial identity differs in its trait impact from the average impact of its component single-SIRE categories. We examine the role of additional explanatory variables in the admixture regression and their interpretation with and without orthogonalization with respect to the core explanatory variables. We generalize the linear admixture regression specification to a partially linear semiparametric form.

We apply our methodology to neuropsychological performance data from the Adolescent Brain Cognitive Development database. Neuropsychological performance is one of the most complex traits to which admixture regression analysis has been applied. Our findings corroborate existing evidence that genetic variation plays a statistically significant role in explaining neuropsychological performance differences across racial identities (Lasker, Pesta, Fuerst, & Kirkegaard, 2019). Using our new test statistic, we find that some multi-racial categories have an identifiably distinct impact relative to their component categories. We find that orthogonalization of additional variables can substantially change the interpretation of the core coefficients in the admixture regression. Our analysis also indicates (although not conclusively) that a partially linear semiparametric specification potentially adds empirical value.

1.1. A Statistical Framework for Admixture Regression Tests of Trait Variation

1.1.1. Variable Definitions

We assume that the database consists of n individuals indexed by $i = 1,\ldots, n$ who have each self-identified their racial or ethnic group membership(s), recorded a score on a behavioral trait, s_i, and provided a personal DNA sample. The k racial or ethnic group self-identification choices are captured by a matrix of zero-one dummy variables $SIRE_{ij}$, $i = 1,\ldots, n; j = 1,\ldots, k$. We assume that every individual has self-identified as belonging to at least one and possibly more of the k groups.

We assume that a set of m geographic ancestries covered in the study have been chosen, such as African, European, Amerindian, South Asian, and East Asian, indexed by $h = 1,\ldots, m$. The genotyped DNA samples are carefully decomposed into admixture proportions of geographic ancestry, as discussed in Section 4 below. For each individual, the ancestry proportions across the chosen geographic ancestries, A_{ih}, $i=1,\ldots, n; h=1,\ldots, m$, are all between zero and one, and they sum to one for each i.

In most applications of admixture regression, individuals' racial or ethnic group identities will have statistical relationships with individuals' genetically identified geographic ancestries and also with the observed trait s_i. The objective of admixture regression is to decompose trait variation into linear components due to genetic ancestries and linear components due to racial/ethnic group-related effects.

1.1.2. Ancestry Proportions as a Statistical Proxy for Ancestry-Linked Genetic Trait Variation

Admixture regression is an indirect method of analyzing group-related trait variation. In this subsection, we provide a foundation for admixture regression by considering a more direct, but empirically much more challenging, alternative approach based on a linear polygenic index model. We show that the admixture regression model can be viewed as a statistically feasible simplification of this linear polygenic index model, in which proportional ancestries serve as statistical proxies for ancestry-related genetic differences.

The human genetic code contains a very large number of genetic variants (the alleles on the genome which vary between individuals) called single nucleotide polymorphisms or SNPs. Consider hypothetically a complete list of all genetic variants with any impact on variation in the observed trait. Assign a value of 0, 1 or 2 to each SNP for individual *i* depending upon the number of minor alleles for that SNP. Let SNP_{iz} $i = 1,\ldots, n$; denote the number of minor alleles on the z^{th} SNP of the i^{th} individual in the sample.

The biochemical process linking human genetic variation to behavioral trait variation is highly complex, and scientific understanding of the full biochemical process is very limited. Genome-wide association studies (GWAS) have made slow but steady progress in statistically modeling these linkages, although precise biochemical linkages are beyond the contemporary scientific frontier for most behavioral traits. A standard, admittedly highly simplified, model of the gene variation/trait variation nexus is the linear polygenic index model, in which the genetic component of a trait is a simple linear function of a relevant subset of the individual's genetic variants. The linear polygenic index model has been applied to a wide range of medical and behavioral traits including body mass index (Yengo et al., 2018), neuroticism (Nagel et al., 2018), depression susceptibility (Wray et al., 2018), suicidal ideation (Mullins et al., 2014), schizophrenia (Mistry, Harrison, Smith, Escott-Price, & Zammit, 2018), educational attainment (Lee et al., 2018), neuropsychological test performance (Savage et al., 2018), and risk-taking (Clifton et al. 2018). The linear admixture regression model can be derived elegantly by invoking this standard linear polygenic index model, and hence we impose it, in order to provide a statistical underpinning for our admixture regression model.

Let p_i denote the genetic potential of individual *i* regarding the observable trait s_i. We assume that p_i is a linear function of a large number of genetic variants SNP_{iz} with associated linear coefficients β_z and constant term c_1:

$$p_i = c_1 + \textstyle\sum_z \beta_z SNP_{zi}; i = 1, \ldots, n. \qquad (1)$$

The key difference in the admixture regression methodology compared to GWAS is that there is no attempt to estimate the linear polygenic index (1). Rather, admixture regression uses the natural experiment of subpopulation mixing to infer differences in the conditional expected value of (1) arising from differences in the frequency distribution of genetic variants across ancestries. The assumed linearity of the polygenic index model (1), together with an assumption of random mating across ancestral populations, allows us to derive a linear regression model using admixture as a statistical proxy variable for the conditional expected value of p_i.

The frequency distributions of many SNPs depend notably upon geographic ancestries. Consider a hypothetical individual with single-origin ancestry *h*, that is, an individual with $A_h = 1$. Note that this also implies that $A_{h'} = 0$ for all $h' \neq h$ since the ancestral proportions are non-negative and sum to one. Consider the expected value of *p* conditional on an individual having this single-origin ancestry. The expectation of (1) using a single-origin frequency distribution for each SNP_z defines the average genetic trait potential of a single-origin ancestry:

$$E[p|A_h = 1] = c_1 + \sum_z \beta_z E[SNP_z|A_h = 1], h = 1, \ldots, m. \quad (2)$$

In admixture regression there is no attempt to measure (2) directly, but instead differences between (2) across $h = 1,\ldots, m$ will be inferred indirectly using regression methods.

A key assumption of the admixture regression model is that admixture arises from recent random mating between the previously geographically isolated ancestral groups. Assuming recent random mating between ancestral lines, it follows from the fundamental processes of sexual reproduction that the expected value of any SNP for an admixed individual is the convex combination of the single-origin expected values, with linear coefficients equal to the individual's admixture proportions. (The relationship between the multivariate distributions of the SNPs is more complicated, but the multivariate distributions do not impact the expected trait given the linear polygenic index assumption.)

We assume that mating across geographic ancestries is recent and random, and therefore in particular that the univariate frequency distribution of each SNP for any individual is the convex combination of the single-origin frequency distributions:

$$E[SNP_{iz}|A_{i\cdot}] = \sum_{h=1}^{m} A_{ih} E[SNP_z|A_h = 1] \quad (3)$$

The linearity of genetic potential in the SNPs (1) and the random mixing assumption (3) imply that expected genetic potential of an admixed individual is a convex combination of the individual's admixture proportions. Taking the expectation of (1) using (2) and (3) the conditional expected value of genetic potential for an individual with admixture proportions $A_{i\cdot}$ is the convex combination of the unobserved values $E[p| A_h = 1]$ with observed linear coefficients A_{ih}:

$$E[p_i|A_{i\cdot}] = \sum_{h=1}^{m} A_{ih} E[SNP_z|A_h = 1] \quad (4)$$

Equation (4) is a fundamental identification condition for the admixture regression methodology. As we discuss below, it allows differences between the single-origin expected values of genetic potential, $E[p| A_h = 1]$, $h = 1,\ldots, m$, to be inferred by regression methods.

1.1.3. Adjusting for Ancestry-Related Environmental Influences on the Trait

Define the environmental component of the trait, e_i, as the observed trait minus genetic potential:

$$s_i = p_i + e_i, \quad (5)$$

where e_i is defined as all trait variation not captured by p_i. Equation (5) is only definitional; later we will impose various conditions on e_i to enable statistical identification of the model. Define $\tilde{p}_i$ as the genetic component of the trait for each i which is not explained by ancestry proportions: $\tilde{p}_i = p_i - E[p_i|A_i]$; by simple substitution into (5) this gives:

$$s_i = c_1 + \sum_{h=1}^{m} A_{ih} E[SNP_z|A_h = 1] + \tilde{p}_i + e_i \quad (6)$$

Recall that the ancestry proportions sum to one for each i, so that one term in the sum in (6) is redundant for the purposes of creating a regression model. Dropping the $h=1$ term from the sum and rewriting (6) without the redundant term gives:

$$s_i = c_2 + \sum_{h=2}^{m} A_{ih}b_{Ah} + \tilde{p} + e_i \tag{7}$$

where $b_{Ah} = (\mathrm{E}[p|A_h = 1] - \mathrm{E}[p|A_1 = 1])$; $h = 2,\ldots, m$, and $c_2 = c_1 + \mathrm{E}[p|A_1 = 1]$.
Equation (7) is still not well-specified as a regression model since the error term $\tilde{p}_i + e_i$ will not be mean zero conditional on A_i, due to racial and ethnic group-related effects in e_i. In order to transform (7) into a regression model it is necessary to add explanatory terms to the regression model to remove the expected value of e_i conditional on A_i. This is accomplished by assuming that the differences in e_i conditional on A_i are dependent on the group self-identification choices, but otherwise not dependent upon admixture proportions.

For expositional simplicity, in this subsection we assume that every individual included in the sample has self-identified as belonging to exactly one from the pre-specified set of k racial or ethnic groups. In this case, the $n \times k$ matrix of racial/ethnic group explanatory variables used in the admixture regression, denoted G, is simply set equal to the SIRE matrix: $G_{ij} = SIRE_{ij}$ for $i = 1,\ldots, n$; $j = 1,\ldots, k$. Multi-racial individuals (those who have self-identified as belonging to two or more groups) will be introduced into the analysis in the next subsection.

We assume that after adjusting for the influence of the group identifiers G_{ij}, the remaining error term in (7) is independent of the ancestry proportions:

$$e_i = c_3 + \sum_{j=2}^{k} b_{Gj}G_{ij} + \tilde{e}_i, \tag{8}$$

where b_{Gj} captures the environmental component associated with membership in group j relative to the reference group $j = 1$, and $\tilde{e}_i$ is assumed to be independent of $A_{i.}$, $G_{i.}$ and $\tilde{p}_i$, and c_3 is a constant term. Combining (7) and (8) produces the key linear admixture regression specification:

$$s_i = c_4 + \sum_{j=2}^{k} b_{Gj}G_{ij} + \sum_{h=2}^{m} A_{ih}b_{Ah} + \varepsilon_i \tag{9}$$

where $\varepsilon_i = \tilde{e}_i + \tilde{p}_i$ and c_4 is a constant term. Note that ε_i has zero mean and constant variance and is independent of $A_{i.}$ and $G_{i.}$. Equation (9) is a well-specified linear regression model.

In many applications, the analyst also has information on the sampling substructure of the data, such as its division into site-specific subsamples. In this case, a linear mixed effects model can be used for estimating (9) rather than ordinary least squares. This involves partially decomposing the residual term ε_i in (9) into linear random effects components linked to data collection site identifiers and/or other subsample identifiers, see Heeringa and Berglund (2020).

1.1.4. Adding Multi-Racial Individuals to the Regression

Recall that $SIRE$ is the n × k matrix of race/ethnicity self-identifications. A key assumption of the admixture regression technique is that the environmental influences associated with racial/ethnic group membership are captured by these group membership self-identification choices. Many individuals self-identify as belonging to two or more racial or ethnic groups, and the group variables used in the regression must be adapted to this reality. In the context of our statistical framework, there are essentially three approaches: evenly splitting the individual's affiliation across their chosen groups, creating a new group for one or more particular multi-racial combinations, or deleting particular multi-racial observations where neither of the other two approaches seem appropriate.

We now allow that some individuals choose more than one category, so that the dummy variable $SIRE_{ij}$ can have more than one unit entry for some i. The simplest regression specification in this case is to assume that the group environment faced by a multi-racial individual is the average of the component group environments:

$$G_{ij} = SIRE_{ij}/\left(\sum_{j*=1}^{k} SIRE_{ij*}\right) \text{ for all } i,j. \tag{10}$$

Although (10) is a reasonable specification, it is restrictive. It is possible to replace (10) with a more general specification at some loss of parsimony. Suppose that we are concerned about imposing the restrictive condition (10) for some common multi-racial choice (such as, for example, Black-White biracial in a US dataset). Let V_1 denote a k–vector with ones for the included race/ethnicity groups in this particular multi-racial combination and zeros elsewhere. We can supplement (10) by adding a $k + 1^{st}$ group and using a different rule for this subset of multi-racials:

$$\begin{aligned} G_{ij} &= 0 \; for \; j = 1, \dots, k \; if \; SIRE_{i\cdot} = V_1 \\ G_{ik+1} &= 1 \; if \; SIRE_{i\cdot} = V_1 \\ G_{ik+1} &= 0 \; if \; SIRE_{i\cdot} \neq V_1 \end{aligned} \tag{11}$$

where $SIRE_{i\cdot} = V_1$ denotes vector equality between these two k–vectors. There are now $k + 1$ groups: the originally specified $SIRE$ groups and a new group for the selected multiracial combination. G becomes a $n \times (k+1)$ matrix, and the regression (9) described in the previous subsection applies exactly as before but with one extra dimension to G. Any small number of defined multi-racial groups can be appended in this way. The only change to the regression methodology is that G becomes a $n \times k^*$ matrix (with an associated increase in the set of estimated parameters) where $k^* - k$ is the number of multiracial combinations added as new categories.

It is not feasible to use rule (11) for all race/ethnicity choice combinations due to lack of parsimony; there are $2^k - k$ potential multi-racial combinations and each one added requires an additional parameter in the regression. It can only be used for the common multi-racial choices where there is sufficient data of that combination in the sample. For all others, it is necessary to stick with the restrictive assumption (10) or drop the observations from the sample. This will be illustrated in the empirical application in Section 5.

Once a regression model is estimated using (11), it is possible to test the accuracy of restrictive assumption (10) for that multi-racial group. The restrictive assumption implicit in (10) requires that the average of the coefficients of the components equals the added-group coefficient in the unrestricted model:

$$\frac{1}{\#j*}\sum_{j*} b_{Gj*} = b_{G,k+1} \tag{12}$$

where $\#j^*$ denotes the number of components in the multiracial category (typically either two or three) and the sum runs over these elements only. This is a linear restriction on the vector of coefficients, or multiple linear restrictions for $k^* - k$ greater than one, which can be tested with a t-test (for each group coefficient singly) or a Wald test for all them, as detailed below.

Let $\hat{b}$ denote the $(m + k^* - 1)$–vector of all the coefficients in the admixture regression (9): $\hat{b} = [\hat{c}_4, \hat{b}_G, \hat{b}_A]$ and let $\widehat{Cov}_b$ denote the estimated $(m + k^* - 1) \times (m + k^* - 1)$–covariance matrix of these estimates.

First consider the case $k^* - k = 1$. Let R denote the $(m + k^* - 1)$–vector expressing restriction (12) imposed on b. For example, if the group combination consists of individuals who choose all three of the first, second, and third SIRE categories (recalling that the first SIRE category is not included in the regression) the restriction vector is: $R = [0, -\frac{1}{3}, -\frac{1}{3}, 0, .., 0, 1, 0, .., 0]$ where the 1 is element k^* in the vector. Any other restriction of type (12) is easily stated in this way. In the case of one group, this gives rise to a standard *t*-test of the one coefficient restriction, and in particular:

$$\frac{(\hat{b}'R)}{R'\widehat{Cov}_b R} \sim t(n - m - k^* + 1) \qquad (13)$$

For the case $k^* - k > 1$ it is possible to test each multi-racial group equality individually as above using (13) or perform a joint Wald test on all of them. Let R denote the $(m + k^* - 1) \times (k^* - k)$– matrix of all the linear restrictions, giving the standard Wald test:

$$(\hat{b}'R)(R'(\widehat{Cov}_b^{-1})R)(R'\hat{b}) \sim \chi^2(k^* - k) \qquad (14)$$

where ~ denotes the approximate distribution for large *n*. In the case of estimation by linear mixed effects modeling, both test statistics (13) and (14) are large–*n* asymptotic distributions rather than exact finite-sample distributions, but they remain valid tests.

1.2. Extensions of the Linear Admixture Regression Model

1.2.1. Variable Definitions

It is straightforward to include additional explanatory variables in the admixture regression model. Let $x_{i1}, x_{i2}, \ldots, x_{il}$ denote a set of explanatory variables that help to linearly explain the trait along with the ancestry proportions and group identities. We modify specification (9) to include these:

$$s_i = c_5 + \sum_{j=1}^{k} b_{Gj} G_{ij} + \sum_{h=2}^{m} b_{Ah} A_{ih} \sum_{d=1}^{l} b_{xd} x_{id} + \varepsilon_i \qquad (15)$$

and keep all the other assumptions as before. The estimation theory for (15) is essentially identical to that of (9) as discussed above.

In some cases, the admixture regression model with additional explanatory variables (15) can be made more useful and informative by orthogonal rotation of one or more of the explanatory variables, in order to aggregate the full effects of proportional ancestries and group identities into their associated coefficients. To understand why such an orthogonal rotation might be useful, consider the hypothetical case of an admixture regression model of Body Mass Index (BMI) in which waist measurement is one of the explanatory variables. Waist measurement has such strong explanatory power for BMI that its presence in an admixture regression model like (15) will diminish the direct explanatory power of proportional ancestries and group identities; their total impact will be partly hidden within the waist measurement variable. This can be remedied by orthogonalizing the waist measurement variable with respect to the proportional ancestry and group identity variables before estimating the admixture regression, as explained next.

Suppose that variable x_1 in (15) has strong explanatory power for s and substantial correlation with proportional ancestry and/or group identity variables, and therefore the analyst wishes to orthogonalize it with respect to G_{ij} and A_{ih}, $j = 2, \ldots, k$; $h = 2, \ldots, m$. In a first step, the analyst can perform a simple least square regression decomposition of x_1 into the component linearly explained by these variables, and the residual, orthogonal component x_1^o :

$$x_1 = \hat{c}_6 + \sum_{j=2}^{k} \hat{b}_{Gj} G_{ij} + \sum_{h=2}^{m} \hat{b}_{Ah} A_{ih} + x_1^o \qquad (16)$$

Since all the explanatory variables are deterministic (that is, conditionally fixed variables rather than random variables in the regression model), this orthogonalization step is interpreted as a matrix transformation of fixed vectors and does not alter any statistical assumptions of the main regression model. It merely serves to linearly rotate the deterministic explanatory variables used in the actual, second-stage, admixture regression. Replacing x_1 with x_1^o in (15) changes the interpretation of the coefficients $\widehat{b}_{Gj}$ and $\widehat{b}_{Ah}$, $j = 2,\ldots, k$; $h = 2,\ldots, m$ since they now include the G_{ij} and A_{ih} related explanatory power from x_1. An illustrative example will be provided in Section 5 below.

1.2.2. A Semiparametric Extension of the Admixture Regression Model

The linear dependence of the trait on admixture proportions in our regression model is in part an artefact of the assumption of a linear polygenic index (1). It is possible to weaken this linearity assumption using nonparametric regression methods. We replace the restrictive assumption of a linear polygenic index (1) with a very general description of genetic potential as a function of the full vector of genetic variants: $p_i = p(SNP_i)$ and instead of linearity as in (1) only require smoothness conditions on the conditional expectation of $p(\cdot)$ as a function of the ancestral proportions vector, as delineated below.

As in earlier subsections, we consider p_i as a stochastic function of the ancestral proportions vector $A_{i\cdot}$, but now without imposing the strict linearity (4) arising from the linear polygenic index assumption: $f(A_{i\cdot}) = E[p(SNP_{i\cdot})|A_{i\cdot}]$. Define the unexplained component of p_i as before; $\tilde{p}_i = p_i - f(A_{i\cdot})$ and we assume that $\tilde{p}_i = N(0, \sigma_p^2)$ and independent of $A_{i.}$ and $G_{i.}$. We impose the same assumptions on ε_i as in Section 2, giving:

$$s_i = \textstyle\sum_{j=1}^{k} b_{Gj} G_{ij} + f(A_{i\cdot}) + \varepsilon_i \tag{17}$$

where ε_i is assumed to be normally distributed with mean zero and constant variance and independent of $A_{i.}$ and $G_{i.}$. This equation (17) is a partially linear nonparametric regression model, see, e.g. Li and Racine (2007).

This model can be consistently estimated using the three-step procedure of Robinson (1988). We will impose Condition 7.1 from Li and Racine (2007) in order to justify this procedure within our framework (see the Technical Appendix for details).

For the case $m > 2$ the general specification (17) suffers from the curse of dimensionality and is unlikely to be estimable on moderate-sized datasets. A more restrictive specification is needed to give the model sufficient parsimony for estimation. One reasonable specification choice is to restrict the nonlinearity in the impact of ancestries on the trait to a single ancestral category, which we assume is ancestry category 2, giving rise to the specification:

$$s_i = \textstyle\sum_{j=2}^{k} b_{Gj} G_{ij} + f_2(A_{i2}) + \sum_{h=3}^{m} b_{Ah} A_{ih} + \varepsilon_i, \tag{18}$$

and we will now rely on this more restrictive specification throughout the remainder of this subsection. We assume that the unconditional density $\Pr(A_2)$ is continuous and strictly positive everywhere on the [0, 1] interval. Let $\widehat{\Pr}\,(A_{i2})$ denote the nonparametrically estimated unconditional density of A_{i2}:

$$\widehat{\Pr}\,(A_{i2}) = \tfrac{1}{n} \textstyle\sum_{i'=1}^{n} k(A_{i'2} - A_{i2}) \tag{19}$$

where *k*(•) is a kernel weighting function. In our empirical application in Section 5 we use the Gaussian kernel weighting function.

In the first step of the Robinson procedure, the conditional means of the dependent variable and linear-component explanatory variables are estimated nonparametrically as functions of the nonparametric-component explanatory variable, A_{i2}:

$$\hat{f}_0(A_{i2}) \approx E[s_i|A_{i2}]$$
$$\hat{f}_{Gj}(A_{i2}) \approx E\left[G_{ij}\middle|A_{i2}\right]; j = 2, \dots, k$$

and

$$\hat{f}_{Ah}(A_{i2}) \approx E[A_{ih}|A_{i2}]; h = 3, \dots, m$$

that is:

$$\hat{f}_0(A_{i2}) = \frac{1}{n}\sum_{i'=1}^{n} s_{i'}k(A_{i'2} - A_{i2})/(\widehat{\Pr}(A_{i2}), j = 2, \dots, k$$
$$\hat{f}_{Gj}(A_{i2}) = \frac{1}{n}\sum_{i'=1}^{n} G_{i'j}k(A_{i'2} - A_{i2})/(\widehat{\Pr}(A_{i2}), j = 2, \dots, k$$

and:

$$\hat{f}_{Ah}(A_{i2}) = \frac{1}{n}\sum_{i'=1}^{n} A_{i'h}k(A_{i'2} - A_{i2})/(\widehat{\Pr}(A_{i2}), h = 3, \dots, m$$

In the second step, the linear parameters of the model (17) are estimated by ordinary least squares, replacing the dependent variable and linear-component explanatory variables with the deviations from their conditional mean functions: $\left(\hat{b}_G, \hat{b}_A\right) = (X'X)^{-1}X'y$ where:

$$y_i = s_i - \hat{f}_0(A_{i2})$$
$$X_{ij} = G_{ij} - \hat{f}_{Gj}(A_{i2}); j = 2, \dots, k,$$
$$X_{ih} = A_{ih} - \hat{f}_{Ah}(A_{i2}); h = 3, \dots, m.$$

Note that $\left(\hat{b}_G, \hat{b}_A\right)$ is a $(k + m - 3)$–vector and X is a $n \times (k + m - 3)$–matrix where the index first runs from 2 to *k* over *j* and then from 3 to *m* over *h*.

In the third step, the nonparametric component of the model is estimated by subtracting the predicted linear component from both sides of (17) and then applying standard nonparametric regression:

$$y_i^* = s_i - \left(\sum_{j=2}^{k} \hat{b}_{Gj}G_{ij} + \sum_{h=3}^{m} \hat{b}_{Ah}A_{ih}\right); i = 1, \dots, n$$

and then:

$$f(A_{i2}) = \frac{1}{c_i}\sum_{i'=1}^{n}(A_{i'2} - A_{i2})y_{i'}^*,$$

where:

$$c_i = \sum_{i'=1}^{n} k(A_{i'2} - A_{i2}).$$

The partially linear nonparametric approach to admixture regression is more empirically challenging than the linear specification. Proper implementation of the technique involves a trade-off between parsimony, the generality of the specification used, and the distributional features of the available data. An example of (18) will be estimated in Section 5 below.

2. Materials and Methods

The Adolescent Brain and Cognitive Development (ABCD) study is the largest long-term study of brain development and child health in the United States, testing 11,000 children ages 9-10 at 21 testing sites; see Karcher and Barch (2021) for an overview.

Our sample consists of age and gender-adjusted scores and genotyped DNA samples of the 9972 children in the ABCD study who met our sample selection criteria, along with questionnaire responses of their parent(s)/guardian(s). The dependent variable in our model is the composite neuropsychological performance score based on the NIH Toolbox (NIHTBX) neurocognitive battery provided in the ABCD database; this consists of tasks measuring attention, episodic memory, language abilities, executive function, processing speed, and working memory. Age-corrected composite scores, based on the seven tasks, were provided by ABCD. We regressed out sex from these age-corrected composite scores. The residuals were then standardized and serve as the dependent variable in our empirical analysis in Section 5 below. Our core explanatory variables are seven SIRE variables, White, Black, Hispanic, Native American, East Asian, South Asian, and Other (and including multiple SIRE choices from among these) and five genetic ancestry proportions of European, African, Amerindian, East Asian and South Asian background obtained from the genotyped DNA samples.

Children whose parent(s)/guardian(s) identified the child as belonging to Pacific Islander racial groups were excluded from our analyses owing to a lack of corresponding ancestry category in our chosen five categories. The ABCD Version 3 database provides 516,598 genotyped SNP variants for each individual's DNA sample. After quality control, filtering, and pruning we were left with 99,642 SNP variants to determine the five ancestry proportions, employing the Admixture 1.3 software package (Alexander, Shringarpure, Novembre, & Lange, 2015). The ABCD database includes site identifiers for the data collection sites (in most 268 cases, elementary schools) and family household identifiers (identifying multiple individuals in the sample from the same family household, usually twins). As recommended by Heeringa and Berglund (2020) for regression analysis using the ABCD database, we include random effects in our regression models to account for any site-specific and family-specific error correlation. We use the linear regression mixed effects estimate routine *lme*r from the R programming language library, see Bates et al. (2015). The one exception is regression Model 3 (see below) in which we estimate a semiparametric partially linear model. In that case, we use the *npplr* routine in the R programming subroutine library *NP,* written and maintained by Hayfield and Racine (2020), and do not correct for site-specific and family-specific error correlation.

See the Supplemental Materials for more detailed description of the ABCD database, our sample selection procedure, and the construction of the variables that we use.

3. Results

In this section, we apply our admixture regression techniques to neuropsychological performance using the ABCD database. Table 1 shows means and standard deviations of the regression dependent variable on data subsets sorted by SIRE choice. On the full sample, by construction, the dependent variable has a mean of zero and standard deviation of one. There is considerable dispersion in the subsample means sorted by SIRE; for example, the means differ by 1.02 standard deviations (using the full-sample standard deviation for simplicity) between two of the largest SIRE categories shown, White-only SIRE and Black-only SIRE. The considerable variation in means for SIRE-based subsamples provides an initial justification for performing admixture regression analysis. This is a table of descriptive statistics; the standard errors shown are not appropriate for formal hypothesis testing since there is no adjustment for potential site-linked and family-linked correlations, particularly relevant in the case of the smaller subsample categories.

Table 1. Neuropsychological Performance Scores Sorted by Self-identified Race or Ethnicity (SIRE)

	Individuals with Single-SIRE Identities						
Within-Category Means and Standard Deviations	White Only	Black Only	Native American Only	East Asian Only	South Asian Only	Other Only	All Hispanic
Mean	0.25	-0.77	-0.42	0.57	0.45	-.22	-.23
Standard Deviation	0.92	0.87	0.79	1.02	1.02	1.12	0.96
Number of Observations	5593	1434	31	107	43	97	1869
	Individuals with Selected Multiple-SIRE Identities						
	Black-White	Hispanic-White	Native American -White	East Asian – White	South Asian - White	Hispanic – Black	Hispanic - Other
Mean	-0.13	-0.19	0.01	0.58	0.83	-0.34	-0.45
Standard Deviation	0.95	0.96	0.86	0.98	0.84	0.93	0.92
Number of Observations	302	1171	131	249	40	84	411

Notes: The table shows means and standard deviations of neuropsychological performance scores for individuals sorted into categories by self-identified race and ethnicity (SIRE). The categories used are: individuals who choose only one of the six race categories (White, Black, Native American, East Asian, South Asian, and Other), all individuals selecting Hispanic ethnicity along with any of the six race categories, and individuals choosing the seven most common two-SIRE choices. By construction the mean and standard deviation of the full sample are zero and one; there are 9972 individuals in the full sample.

Table 2 displays empirical results from three specifications of the admixture regression methodology. Recall that one SIRE variable and one ancestry proportion variable must be left out as an identification condition of the admixture regression: we leave out the White SIRE variable and the European ancestry proportion variable. Model 1 uses a linear regression specification and singleton SIRE categories for the group-identity variables G; individuals who choose multiple SIRE categories have G exposures equally divided between the chosen SIRE categories. Three of the four ancestral proportion variables and one of the six group-identity variables have statistically significant coefficients. Model 2 adds a selected set of multiple-SIRE composite categories to the G specification.

We include the seven two-category choices with the largest number of observations in our sample. Any individual with one of these two-category choices has unit exposure to the associated explanatory variable, and no exposure to the weighted single-SIRE variables (see Equation 11 above). The same three of four ancestral proportion variables as in Model 1 are significant in Model 2, with similar coefficients to Model 1. None of the single-SIRE group identity variables is significant. Three of the seven selected two-SIRE group identity variables have significantly different coefficients from that implied by equal weightings of the component single-category coefficients. One of these (Hispanic-Other) has a statistically significant coefficient; the other two are not significantly different from zero, but are significantly different from the value implied by the composite single-category coefficients. Random effects are included in all models except Model 3 to capture any common variation associated with the individual data collection sites in the ABCD study or associated with those households having multiple individuals in the sample.

We use the *lmer* maximum likelihood mixed effects model estimation routine from the *R* language library, Bates, Maechler, and Bolker (2015), for all models except Model 3. See Nakagawa and Schielzeth (2013) for the definition and interpretation of conditional and marginal R^2 in a linear mixed effects model. The marginal R^2 (which does not include the explanatory power associated with site-specific and family-specific random effects) is approximately 0.16 in both model specifications. The African, Amerindian, and East Asian proportional ancestry variables have strong and significant explanatory power in Models 1 and 2. For single-SIRE individuals, the SIRE-based group identity variables are mostly indistinguishable from zero, but some of the multiple-SIRE group variables are significantly different from zero.

Table 2: Admixture Regression Results for Neuropsychological Performance: Linear Specifications with and without Composite Groups and a Partially Linear Semiparametric Specification

	Core Explanatory Variables										
	Intercept	% African	% Amer-indian	% East Asian	% South Asian	Black SIRE	Hispanic SIRE	Native American SIRE	East Asian SIRE	South Asian SIRE	Other SIRE
Model 1	0.30	-1.02	-1.38	0.64	0.49	-0.14	-0.09	-0.14	-0.21	-0.07	-0.20
t-statistic	9.06	-8.34	-11.37	3.19	1.49	-1.41	-1.12	-1.33	-1.21	-0.28	-2.52
Model 2	0.29	-1.03	-1.34	0.66	0.54	-0.15	0.34	-0.19	-0.35	-0.27	-0.18
t-statistic	8.93	-8.23	-10.78	3.25	1.66	-1.46	1.73	-1.37	-1.93	-0.99	-1.76
Model 3	N/A	Figure 3	-1.19	0.69	0.74	-0.13	-0.12	-0.26	-0.14	-0.15	-0.08
t-statistic			-9.95	3.48	2.31	-1.33	-1.45	-2.52	-0.77	-0.60	-0.91
	Multiple-SIRE-Composite Explanatory Variables										
	Black-White SIRE	Hispanic-White SIRE	Native America - White SIRE	East Asian - White SIRE	South Asian - White SIRE	Hispanic - Black SIRE	Hispanic - Other SIRE	Wald Test Statistic	Wald Test *p*-value		
Model 2 [cont.]	0.05	-0.08	-0.10	0.00	0.34	0.04	-0.17				
t-statistic	0.71	-1.79	-1.19	0.04	1.80	0.29	-2.40				
Test 2	2.28	-2.55	-0.07	0.94	1.51	-0.42	-2.01	27.08	0.00		
Conditional R^2	Model 1: 0.550; Model 2: 0.550; Model 3:NA										
Marginal R^2	Model 1: 0.157; Model 2: 0.160; Model 3:NA										

Note: Model 1 uses single-SIRE categories with multiple-SIRE choices allocated evenly across them; Model 2 adds seven multiple-SIRE categories. Model 4 uses semiparametric estimation and single-SIRE categories as in Model 1. Group Component test gives the z-statistic for testing if the multiple-SIRE group coefficient equals the average of the component coefficients; the Wald statistic provides a joint test of all the group component test restrictions.

Model 3 implements a partially linear nonparametric specification. This specification requires that the highlighted ancestry proportion (whose impact is estimated nonparametically) has observations throughout the [0, 1] range. For each of the five ancestry categories, Table 3 gives the number of sample observations of proportional ancestry in decile bins of percent ancestry for each of the five genetic ancestry categories. We use African proportional ancestry as the highlighted variable since it fulfils the requirement for observations throughout the [0, 1] interval, and therefore, partially linear nonparametric estimation is feasible. Figure 1 shows the probability density of African ancestry for the full sample population; Figure 2 shows the density restricted to those individuals having measured African ancestry greater than 0.5%, this provides greater detail in the graph by excluding observations with near-zero ancestry. Interestingly, this density has three local peaks, at approximately 5%, 40% and 80% African ancestry.

Table 3: Number of Observations in Deciles of Proportional Ancestry for Each Ancestry Category

European	Interval	$A_{1i}\leq10\%$	$10\%<A_{1i}\leq20\%$	$20\%<A_{1i}\leq30\%$	$30\%<A_{1i}\leq40\%$	$40\%<A_{1i}\leq50\%$
	Number of Obs.	298	908	425	346	461
	Interval	$50\%<A_{1i}\leq60\%$	$60\%<A_{1i}\leq70\%$	$70\%<A_{1i}\leq80\%$	$80\%<A_{1i}\leq90\%$	$90\%<A_{1i}$
	Number of Obs.	700	406	462	514	5452
African	Interval	$A_{2i}\leq10\%$	$10\%<A_{2i}\leq20\%$	$20\%<A_{2i}\leq30\%$	$30\%<A_{2i}\leq40\%$	$40\%<A_{2i}\leq50\%$
	Number of Obs.	7557	2935	286	125	165
	Interval	$50\%<A_{2i}\leq60\%$	$60\%<A_{2i}\leq70\%$	$70\%<A_{2i}\leq80\%$	$80\%<A_{2i}\leq90\%$	$90\%<A_{2i}$
	Number of Obs.	88	130	406	787	149
Amerindian	Interval	$A_{3i}\leq10\%$	$10\%<A_{3i}\leq20\%$	$20\%<A_{3i}\leq30\%$	$30\%<A_{3i}\leq40\%$	$40\%<A_{3i}\leq50\%$
	Number of Obs.	8364	443	329	301	282
	Interval	$50\%<A_{3i}\leq60\%$	$60\%<A_{3i}\leq70\%$	$70\%<A_{3i}\leq80\%$	$80\%<A_{3i}\leq90\%$	$90\%<A_{3i}$
	Number of Obs.	156	75	18	0	4
East Asian	Interval	$A_{4i}\leq10\%$	$10\%<A_{4i}\leq20\%$	$20\%<A_{4i}\leq30\%$	$30\%<A_{4i}\leq40\%$	$40\%<A_{4i}\leq50\%$
	Number of Obs.	9455	74	84	20	225
	Interval	$50\%<A_{4i}\leq60\%$	$60\%<A_{4i}\leq70\%$	$70\%<A_{4i}\leq80\%$	$80\%<A_{4i}\leq90\%$	$90\%<A_{4i}$
	Number of Obs.	18	4	8	10	74
South Asian	Interval	$A_{5i}\leq10\%$	$10\%<A_{5i}\leq20\%$	$20\%<A_{5i}\leq30\%$	$30\%<A_{5i}\leq40\%$	$40\%<A_{5i}\leq50\%$
	Number of Obs.	9796	55	30	29	17
	Interval	$50\%<A_{5i}\leq60\%$	$60\%<A_{5i}\leq70\%$	$70\%<A_{5i}\leq80\%$	$80\%<A_{5i}\leq90\%$	$90\%<A_{5i}$
	Number of Obs.	4	11	9	21	0

Note: For each of the five geographic ancestries, the table shows the number of the 9972 total observations within each of the deciles of proportional ancestry.

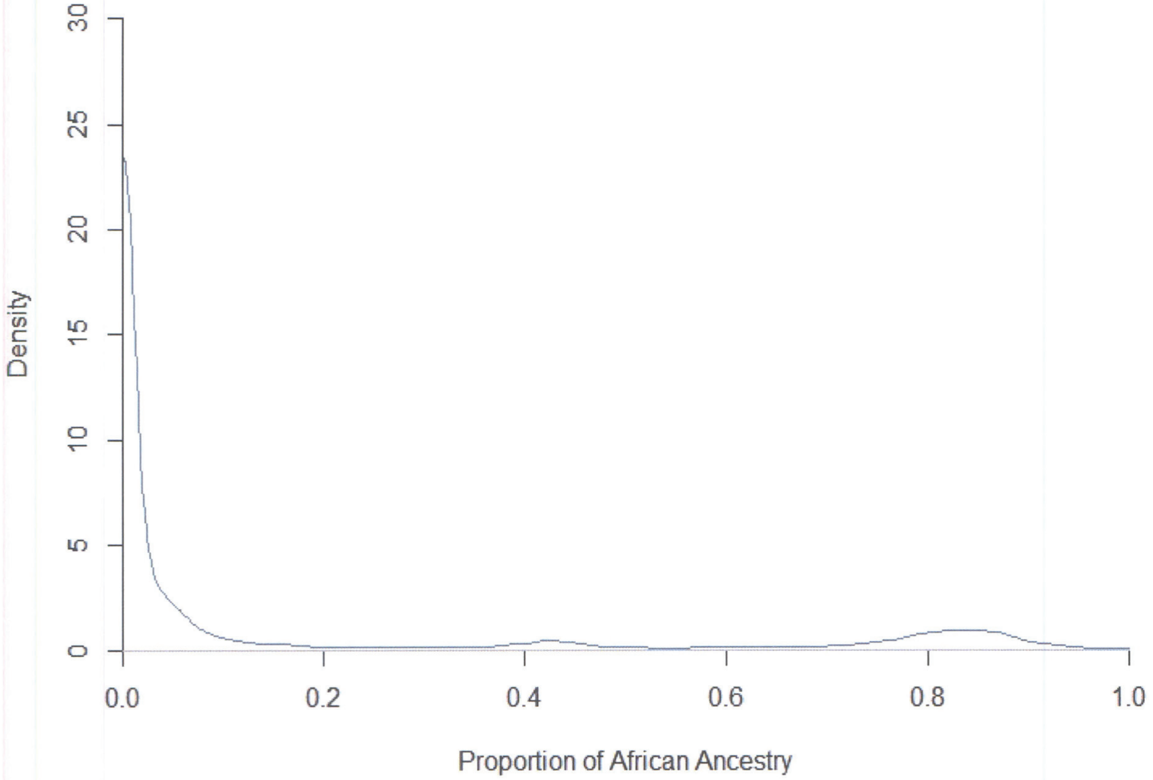


Figure 1: Estimated Density of African Ancestry for the Full Sample

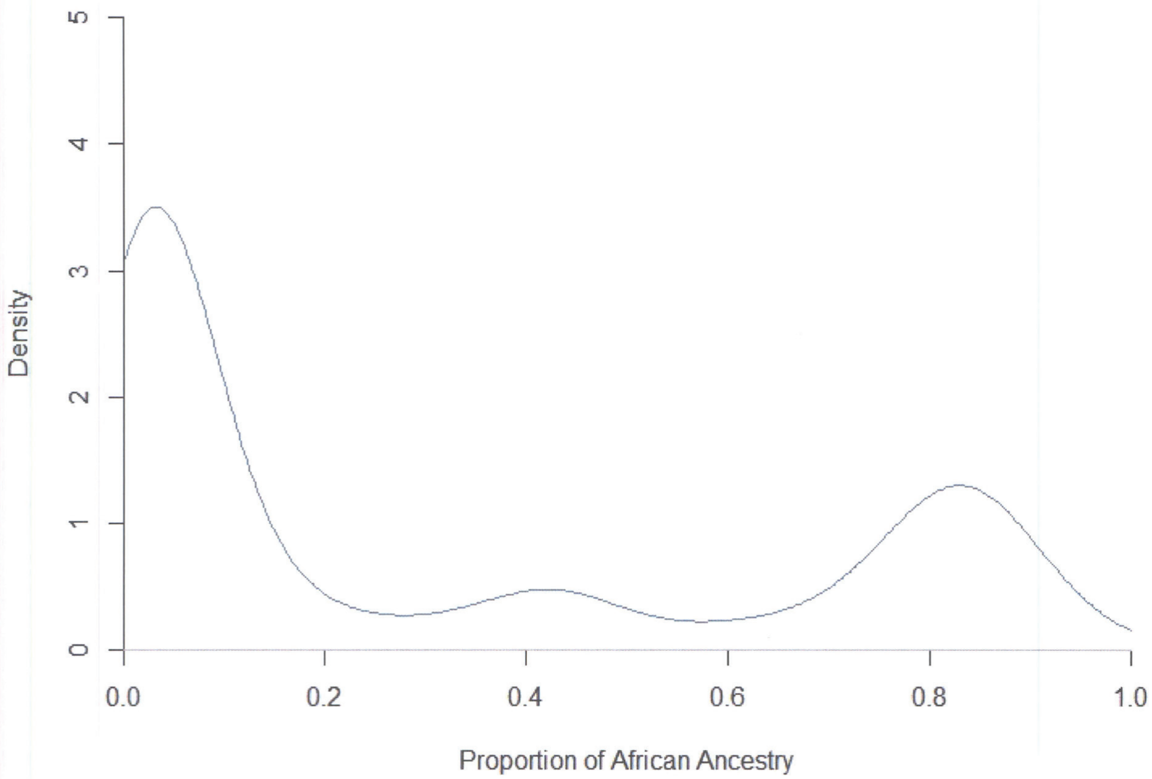


Figure 2: Estimated Density of African Ancestry for a Restricted Sample (Ancestry > 0.5%)

Partially linear semiparametric Model 3 (18) is estimated using the *npplr* routine in the R programming language subroutine library *NP* written and maintained by Hayfield and Racine (2020). We use the simple average SIRE specification of *G* as in Model 1. We use the Gaussian kernel throughout, and all bandwidths are chosen by iterated least-squares cross-validation. The linear coefficient estimates in Model 3 do not differ notably from those in Model 1. Figure 3 displays the nonparametric estimate of the impact of African ancestry on the performance variable along with the corresponding linear impact estimate from Model 1, that is, $\hat{f}(A_2) - \hat{f}(0)$ and $A_2\hat{b}_2$ for $A_2 \in [0, 1]$. There is some graphical evidence for an uptick in the nonlinear gradient for ancestry proportions above 90%. We now briefly examine this further.

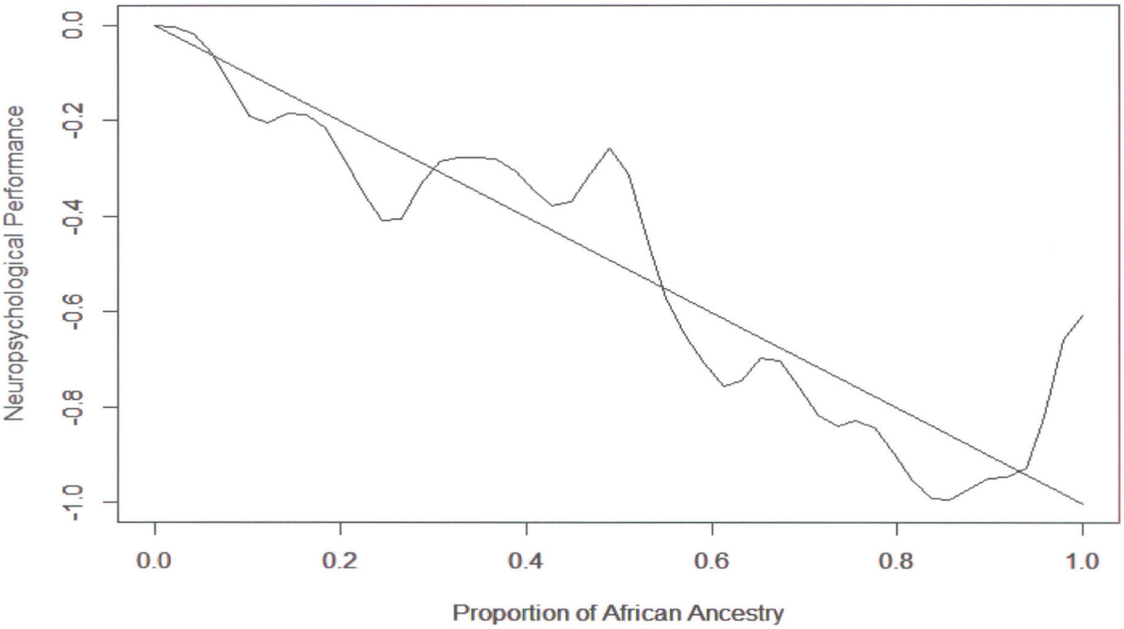


Figure 3: Linear and Nonlinear Gradients Measuring the Impact of African Ancestry

Model 3 does not capture the efficiency gain and test statistic bias reduction from the mixed effects modeling used in the estimation of the other models. Figure 3 of Model 3 is estimated in the second stage of a two-stage semiparametric estimation process and this weakens its empirical reliability. To examine more carefully the graphical pattern observed in Figure 3, but with single-stage estimation and the advantage of mixed effects modeling, we estimate a piecewise linear specification for $A_{i2} \geq 0.9$. This was chosen in order to mimic the observed nonlinear uptick seen in Figure 3 within a linear regression functional form. Recall that African ancestry proportion is ancestry variable 2, giving the formulation:

$$s_i = c + \sum_{j=2}^{7} b_{Gj} G_{ij} + \sum_{h=2}^{5} b_{Ah} A_{ih} + b^{kink} A_{i2} D[A_{i2} \geq 0.9] + \varepsilon_i, \qquad (20)$$

where $D[\bullet]$ is a zero-one dummy variable and b^{kink} is the added coefficient. The results are shown as Models 4 and 5 in Table 4. In Model 4 we use the simple average SIRE specification of *G* as in Model 1; Model 5 adds the same seven two-SIRE combination groups as in Model 2. The coefficient b^{kink} is significantly positive in one of the two models; the significance of this finding must be treated with caution since the particular kink specification (20) is based on examination of Figure 3 using the same data.

Table 4: Piecewise Linear Admixture Regression Results with and without Composite Groups

	Core Explanatory Variables										
	Intercept	% African	% Amer-indian	% East Asian	% South Asian	Black SIRE	Hispanic SIRE	Native American SIRE	East Asian SIRE	South Asian SIRE	Other SIRE
Model 4	0.30	-1.08	-1.39	0.64	0.48	-0.11	-0.08	-0.14	-0.20	-0.07	-0.19
t-statistic	9.11	-8.53	-11.44	3.15	1.46	-1.11	-0.97	-1.26	-1.17	-0.26	-2.41
Model 5	0.30	-1.09	-1.34	0.65	0.53	-0.12	0.35	-0.17	-0.34	-0.26	-0.17
t-statistic	8.99	-8.48	-10.84	3.20	1.62	-1.12	1.78	-1.27	-1.87	-0.96	-1.62
	Piecewise Linear Variable	Multiple-SIRE-Composite Explanatory Variables									
	$D[A_2 \geq 0.9]A_2$	Black-White SIRE	Hispanic-White SIRE	Native America - White SIRE	East Asian - White SIRE	South Asian - White SIRE	Hispanic - Black SIRE	Hispanic - Other SIRE	Wald Test Statistic	Wald Test *p*-value	
Model 4 [cont.]	0.16										
t-statistic	1.81										
Model 5 [cont.]	0.18	0.08	-0.07	-0.10	0.01	0.34	0.07	-0.16			
t-statistic	2.04	1.03	-1.65	-1.18	0.08	1.82	0.57	-2.27			
Test 2		2.42	-2.54	-0.14	0.93	1.51	-0.34	-2.04	27.98	0.00	
Conditional R^2		Model 4: 0.550; Model 5: 0.549									
Marginal R^2		Model 4: 0.157; Model 5: 0.160									

Note: Model 4 uses single-SIRE categories with multiple-SIRE choices allocated evenly across the categories; Model 5 adds seven multiple-SIRE categories. Both models include a kinked-linear explanatory variable for African ancestry above 90\%. Composite group test gives the z-statistic for testing if the multiple-SIRE group coefficient equals the average of the component coefficients; the Wald test provides a joint test of all the composite group test restrictions.

Table 5 and Table 6 add two new variables, US-born child and socioeconomic status (SES), to the admixture regression model. The US-born child variable equals one if the child was born in the US, and zero if born elsewhere. SES is a factor-analytic composite of underlying variables from the ABCD database including neighbourhood SES, subjective SES as determined from a set of questionnaire answers by the parent(s)/guardian(s) of the child on financial adversity, parental/guardian marital status, parental/guardian employment status, the completed level of parental/guardian education, reported neighbourhood safety, and family income. See the Supplementary Materials for a more detailed discussion.

Table 5: Admixture Regression Results Including Raw Social-Economic Status (SES) and US-Born Child Variables

	Core Explanatory Variables										
	Intercept	% African	% Amer-indian	% East Asian	% South Asian	Black SIRE	Hispanic SIRE	Native American SIRE	East Asian SIRE	South Asian SIRE	Other SIRE
Model 6a	0.08	-0.63	-0.82	0.62	0.43	-0.10	-0.03	-0.03	-0.22	-0.11	-0.08
t-statistic	1.29	-5.07	-6.87	3.17	1.36	-0.97	-0.38	-0.33	-1.28	-0.45	-1.00
Model 7a	0.08	-0.65	-0.78	0.63	0.47	-0.10	0.38	-0.05	-0.32	-0.28	-0.04
t-statistic	1.37	-5.13	-6.43	3.20	1.48	-0.97	1.98	-0.39	-1.79	-1.09	-0.45
	Piecewise Linear Variable	Socio-Economic Variables		Multiple-SIRE-Composite Explanatory Variables							
	$D[A_2 \geq 0.9]A_2$	Socio-Economic Status	US Born	Black-White SIRE	Hispanic-White SIRE	Native America - White SIRE	East Asian - White SIRE	South Asian - White SIRE	Hispanic - Black SIRE	Hispanic - Other SIRE	Wald test and *p*-value
Model 6a [cont.]	0.02	0.28	0.10								
t-statistic	0.24	23.73	1.77								
Model 7a [cont.]	0.04	0.28	0.09	0.12	-0.04	-0.06	-0.03	0.28	0.06	-0.08	
t-statistic	0.52	23.71	1.57	1.61	-0.97	-0.72	-0.32	1.57	0.54	-1.14	26.71
Test 2				3.07	-2.43	-0.32	0.68	1.41	-0.57	-2.05	0.00

Table 6: Admixture Regression Results Including Orthogonalized Social-Economic Status (SES) and US-Born Child Variables

	Core Explanatory Variables										
	Intercept	% African	% Amerindian	% East Asian	% South Asian	Black SIRE	Hispanic SIRE	Native American SIRE	East Asian SIRE	South Asian SIRE	Other SIRE
Model 6b	0.20	-1.12	-1.36	0.72	0.62	-0.12	-0.09	-0.24	-0.21	-0.12	-0.17
t-statistic	3.23	-9.11	-11.60	3.66	1.96	-1.20	-1.04	-2.33	-1.25	-0.47	-2.29
Model 7b	0.20	-1.13	-1.31	0.73	0.68	-0.12	0.36	-0.27	-0.35	-0.31	-0.16
t-statistic	3.32	-9.04	-10.98	3.71	2.12	-1.22	1.88	-2.05	-1.97	-1.20	-1.60
	Piecewise Linear Variable	Socio-Economic Variables		Multiple-SIRE-Composite Explanatory Variables							
	$D[A_2 \geq 0.9]A_2$	Socio-Economic Status	US Born	Black-White SIRE	Hispanic-White SIRE	Native America - White SIRE	East Asian - White SIRE	South Asian - White SIRE	Hispanic - Black SIRE	Hispanic - Other SIRE	Wald test and *p*-value
Model 6b [cont.]	0.21	0.28	0.10								
t-statistic	2.39	23.73	1.77								
Model 7b [cont.]	0.23	0.28	0.09	0.07	-0.08	-0.16	0.00	0.32	0.08	-0.15	
t-statistic	2.63	23.71	1.57	1.00	-1.82	-1.98	-0.04	1.76	0.67	-2.22	30.59
Test 2				2.48	-2.72	-0.27	0.92	1.57	-0.28	-2.10	0.00
Conditional R^2	Models 6a,b: 0.546; Models 7a,b: 0.546										
Marginal R^2	Models 6a,b: 0.218; Models 7a,b: 0.220										

Note: Models 6a,b use single-SIRE categories with multiple-SIRE choices allocated evenly among them; Models 7a,b add seven multiple-SIRE categories. All models include a kinked-linear explanatory variable for African ancestry above 90%, a dummy variable for a child born in the US, and a composite variable measuring social-economic status. In Models 6a and 7a the social-economic status variable is in raw form whereas in Models 6b and 7b it is orthogonalized with respect to the other explanatory variables (except US born). Composite group test gives the z-statistic for testing if the multiple-SIRE group coefficient equals the average of the component coefficients; the Wald statistic provides a joint test of all the composite group test restrictions.

Models 6 and 7 are identical to Models 4 and 5 (respectively) from Table 3, except for the addition of these two variables. As discussed in Section 3 above, including additional explanatory variables complicates the interpretation of an admixture regression model in terms of the implied decomposition of trait variation into linear components linked to group identities and components linked to genetic ancestries. The SES variable covaries strongly with both genetic and environmental components of neuropsychological performance scores. To retain the standard interpretability of the admixture regression, it is important to orthogonalize SES with respect to the group identity and ancestry variables before running the regression. For completeness, Models 6 and 7 are shown with and without the orthogonalization of SES; versions a and b of each model in Table 5 and Table 6 respectively. If the purpose of the estimation is to identify the total impact of SES on the trait, the regression with raw SES is more appropriate (version a). For admixture analysis intended to capture the total effects of group identity and genetic ancestry on the trait, orthogonalized SES is more appropriate (version b).

Adding SES to the admixture regression model increases the marginal R^2 from approximately 0.16 to 0.22. If SES is used in its raw form, the coefficients associated with proportional ancestries tend to decrease in magnitude, but the coefficients on African, Amerindian, and East Asian proportional ancestry remain strong and significant. There is no clear and reliable impact on the SIRE-based group-identity coefficients from using SES, in either its raw or orthogonalized form. An example of (18) will be estimated in Section 5 below.

4. Discussion and Limitations

Many behavioral traits covary strongly with racial/ethnic self-identities, but it is often ambiguous whether this covariance reflects environmental causes associated with racial/ethnic identity groups or reflects underlying genetic similarity among group members arising from shared geographic ancestry. Admixture regression relies on the natural experiment of the recent genetic admixture of previously geographically isolated ancestral groups to measure the explanatory power arising from racial/ethnic group identities and that arising from ancestry-based similarities of genetic background. The admixture regression methodology, in various formulations, has been applied to a wide range of medical and behavioral traits, including asthma, obesity, type 2 diabetes, hypertension, neuropsychological performance, and sleep depth.

This paper provided a statistical framework for admixture regression based on the linear polygenic index model of behavioral genetics, and developed refinements and extensions of the methodology within this framework. We provided a simple new test procedure for determining whether multiple-SIRE categories have independent explanatory power not captured by the individual component categories. We considered additional explanatory variables in the admixture regression and their interpretation with and without orthogonalization with respect to core variables. We weakened the linearity assumption and developed a partially linear semiparametric regression specification.

We applied our methodology to neuropsychological performance test data from the Adolescent Brain Cognitive Development database. We confirmed existing findings that genetic variation plays a role in neuropsychological performance differences across self-identified races (Lasker et al., 2019). We found mixed evidence regarding the independent explanatory power of multi-racial identities relative to their component single-race categories. We found that when socioeconomic status (SES) is included as an explanatory variable in the admixture regression, pre-regression orthogonalization of SES has a substantial impact on the measured magnitude of the ancestry proportion coefficients. We found that the proportional ancestry variable associated with African ancestry shows some evidence of nonlinearity in its impact on neuropsychological performance.

The techniques that we proposed for admixture regression studies have broad applicability, but they do have some limitations. We described three approaches to accommodating multiple-identity individuals in admixture regression studies (equal weighting, adding new groups, and deletion of some observations) but none of the three methods are fool-proof in terms of correctly capturing identity-related environmental influences in a parsimonious way. We described how to orthogonalize additional explanatory variables in order to accommodate them in an admixture regression while still capturing the full effect of ancestry-related genetic variation in the ancestry proportions coefficients. A limitation of this orthogonalization procedure is that it does not fundamentally alter the underlying regression being estimated—it merely rotates the estimated coefficients to aid in their interpretation. The partially linear admixture regression method that we describe has the usual limitations of nonparametric and semi-parametric estimation methods. It cannot be applied with complete generality due to the curse of dimensionality, and is data-intensive due to the nonparametric estimation component of the procedure.

The following are available online, at
https://www.biorxiv.org/content/10.1101/2021.05.14.444173v3.supplementary-material,
Supplemental File 2: R Estimation Code Library.

Author contributions: Both authors contributed equally to this study.

Funding: This research received no external funding.

Acknowledgements: Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study held in the NIMH Data Archive (NDA). The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. ABCD consortium investigators designed and implemented the study and or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the research results and interpretations of the authors alone and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Version 2.01 for phenotypic data and Version 3.01 for genotypic data. Additional support for this work was made possible from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680), and Children and Screens: Institute of Digital Media and Child Development Inc.

This study relies exclusively upon archival data from the Adolescent Behavior Cognitive Development (ABCD) database and hence was not subject to institutional review.

Technical Appendix.

In this technical appendix we re-state condition 7.1 from (Racine and Li 2007, p. 224) in the context of our partially linear admixture regression model (18).

We assume that the $(k + m - 1)$–vector of observations (s_i, G_{ij}, A_{ih}) $j = 2,\ldots, k$; $h = 2,\ldots, m$ has an i.i.d. distribution over observations $i = 1,\ldots, n$ and that the conditional mean functions E $[G_{ij}|A_{i2}]$

and E[$A_{ih}|A_{i2}$] are twice differentiable throughout the interior of the domain of A_2, the closed unit interval. Let $m(\bullet)$ denote any of these conditional mean functions or their first or second derivative functions. As in Racine and Li, we impose the following Lipschitz-type smoothness condition on these conditional mean functions and their first and second derivatives:

$$|m(A_2) - m(A_2')| \leq H(z)|A_2 - A_2'|$$

where $H(\bullet)$ is some continuous function such that E[$H(A_2)^2$] is finite. The expectation of $H(A_2)^2$ is over the probability distribution of A_2.

We continue to assume that ε_i is mean-zero normally distributed with constant variance. Since G_{ij} only takes the values of zero and one and A_{ih} is confined to the unit interval, it necessarily follows that both have bounded fourth moments. We assume that $k(\bullet)$ is a bounded second-order kernel.

To formally derive the limiting distribution of the Robinson estimator, it is necessary to define a trimming parameter which ensures that the estimates $\widehat{\Pr}(A_{i2})$ are bounded away from zero. Let t denote a trimming parameter and consider the estimator described in the text but where observations such that $\widehat{\Pr}(A_{i2}) < t$ in (19) are dropped from the subsequent estimation steps. Let ϕ denote the kernel bandwidth for sample size n. Assume that the trimming parameter obeys the following two limiting conditions as $n \to \infty$: $n\phi^2 t^4 \to \infty$ and $nt^{-4}\phi^8 \to 0$.

Under these conditions we have from (Robinson 1988) that:

$$dlim\, \sqrt{n}\left[\left(\hat{b}_{G\cdot}, \hat{b}_A\right) - (b_G, b_A)\right] \sim N(0, \sigma_\varepsilon^2 E[X'X]$$

where the matrix X is defined in the main text of the paper above, in step two of the Robinson procedure.

References

Alexander, D. H., Shringarpure, S. S., Novembre, J., & Lange, K. (2015). Admixture 1.3 software manual. Simon Laboratory, University of Wisconsin, Bioinformatics Programs, Madison, WI.

Bates, D., Maechler, M., & Bolker, B. (2015). Walker., S. Fitting linear mixed-effects models using lme4. J Stat *Softw*, 67(1), 1-48.

Cheng, C. Y., Reich, D., Haiman, C. A., Tandon, A., Patterson, N., Elizabeth, S., ... & Kao, W. L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: a genetic admixture analysis in three US population cohorts. *PloS One*, 7(3), e32840.

Choquet, H., Yin, J., & Jorgenson, E. (2021). Cigarette smoking behaviors and the importance of ethnicity and genetic ancestry. *Translational Psychiatry*, 11(1), 1-10.

Choudhry, S., Burchard, E. G., Borrell, L. N., Tang, H., Gomez, I., Naqvi, M., ... & Risch, N. J. (2006). Ancestry–environment interactions and asthma risk among Puerto Ricans. *American Journal of Respiratory and Critical Care Medicine*, 174(10), 1088-1093.

Clifton, E. A., Perry, J. R., Imamura, F., Lotta, L. A., Brage, S., Forouhi, N. G., ... & Day, F. R. (2018). Genome–wide association study for risk taking propensity indicates shared pathways with body mass index. *Communications Biology*, 1(1), 1-10.

Halder, I., Matthews, K. A., Buysse, D. J., Strollo, P. J., Causer, V., Reis, S. E., & Hall, M. H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep*, 38(8), 1185-1193.

Hayfield, T., & Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5), 1-32.

Heeringa, S. G., & Berglund, P. A. (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. BioRxiv. doi:10.1101/2020.02.10.942011.

Li, Q., & Racine, J. S. (2007). Nonparametric econometrics: theory and practice. Princeton University Press:

Princeton, NJ, USA, 2007.

Karcher, N. R., & Barch, D. M. (2021). The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology*, 46(1), 131-142.

Klimentidis, Y. C., Dulin-Keita, A., Casazza, K., Willig, A. L., Allison, D. B., & Fernandez, J. R. (2012). Genetic admixture, social–behavioural factors and body composition are associated with blood pressure differently by racial–ethnic group among children. *Journal of Human Hypertension*, 26(2), 98-107.

Klimentidis, Y. C., Miller, G. F., & Shriver, M. D. (2009). The relationship between European genetic admixture and body composition among Hispanics and Native Americans. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, 21(3), 377-382.

Lackland, D. T. (2014). Racial differences in hypertension: implications for high blood pressure management. *The American Journal of the Medical Sciences*, 348(2), 135-138.

Lasker, J., Pesta, B. J., Fuerst, J. G., & Kirkegaard, E. O. (2019). Global ancestry and cognitive ability. *Psych*, 1(1), 431-459.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50(8), 1112-1121.

Llibre-Guerra, J. J., Li, Y., Allen, I. E., Llibre-Guerra, J. C., Rodríguez Salgado, A. M., Peñalver, A. I., ... & Llibre-Rodríguez, J. J. (2021). Race, Genetic Admixture, and Cognitive Performance in the Cuban Population. *The Journals of Gerontology: Series A*.

Mistry, S., Harrison, J. R., Smith, D. J., Escott-Price, V., & Zammit, S. (2018). The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: systematic review. *Schizophrenia Research*, 197, 2-8.

Mullins, N., Perroud, N., Uher, R., Butler, A. W., Cohen-Woods, S., Rivera, M., ... & Lewis, C. M. (2014). Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: A genome-wide association and polygenic scoring study. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 165(5), 428-437.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., De Leeuw, C. A., Bryois, J., ... & Posthuma, D. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics*, 50(7), 920-927.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931-954.

Salari, K., Choudhry, S., Tang, H., Naqvi, M., Lind, D., Avila, P. C., ... & Ziv, E. (2005). Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 29(1), 76-86.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., ... & Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, 50(7), 912-919.

Wang, Y., & Beydoun, M. A. (2007). The obesity epidemic in the United States—gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. Epidemiologic Reviews, 29(1), 6-28.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., ... & Viktorin, A. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, 50(5), 668-681.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., ... & GIANT Consortium. (2018). Meta-analysis of genome-wide association studies for height and body mass index in~ 700000 individuals of European ancestry. *Human Molecular Genetics*, 27(20), 3641-3649.

Supplemental File 1. Supplementary material for "Linear and partially linear models of behavioral trait variation using admixture regression" in which the dataset, variables, and methods for the empirical analysis are detailed.

1. **Materials and Methods**

1.1. Dataset

The Adolescent Brain Cognitive Development Study (ABCD) is a collaborative longitudinal project between 21 sites across the US. Its goal is to further research into the psychological and neurobiological basis of development. At baseline, around 11,000 9-10 year old children were sampled, using a probabilistic sampling strategy, from public and private elementary schools and through non-school-based community outreach between 2016 and 2018, with the goal of creating a broadly representative sample of US children of this age. Children who were not fluent in English (or whose parents were not fluent in either English or Spanish) were excluded, along with those with severe medical, neurological, or psychiatric conditions. Informed consent was provided by parents. For phenotypic data, ABCD 2.1 data release was used. For this analysis, we excluded individuals who did not have NIH Toolbox® results, who did not have admixture data, and who were identified as being a Pacific Islander. This left 9972 individuals.

1.2. Variables

1.2.1. Admixture

Subjects were genotyped using Illumina XX, with 516,598 variants directly genotyped and surviving the quality control done by the data provider. We used (updated with) the 3.0 release of the genotypic dataset, which also includes an edition with imputed variants using TOPMED and Eagle 2.4. Because we had very few samples from Pacific Islanders, we excluded these from further analysis to simplify the reference populations needed (*N* = 69). All our work was done on build 38. Files in hg17/37 were lifted to hg38 using liftOver (https://github.com/sritchie73/liftOverPlink) and the GRC chain file at
ftp://ftp.ensembl.org/pub/assembly_mapping/homo_sapiens/ (GRCh37_to_GRCh38.chain.gz).

Before global admixture estimation, we applied quality control using plink 1.9. We used only directly genotyped, bi-allelic, autosomal SNP variants (494,433, 493,196, before and after lifting). We pruned variants for linkage disequilibrium at the 0.1 R^2 level using plink 1.9 (--indep-pairwise 10000 100 0.1), as recommended in the admixture documentation
(https://vcru.wisc.edu/simonlab/bioinformatics/programs/admixture/admixture-manual.pdf).

This variant filtering was done in the reference population dataset to reduce bias from sample representativeness. After pruning, we were left with 99,642 variants. To ensure a reasonable balance in the estimation dataset, we merged the target samples from ABCD, with reference population data for the populations of interest. We desired a k=5 solution (European, Amerindian, African, East Asian, and South Asian), so we merged with relevant samples from 1000 Genomes and from the HGDP. The following populations were excluded: Adygei, Balochi, Bedouin, Bougainville, Brahui, Burusho, Druze, Hazara, Makrani, Mozabite, Palestinian, Papuan, San, Sindhi, Uygur, Yakut. These reference populations were excluded because they were overly admixed or because, in the case of Melanesians and San, the individuals in the ABCD sample lacked significant portions of these ancestries.

Because the estimation sample would still be very skewed towards European ancestry using this joint sample, we used repeated subsetting to achieve balance. Specifically, we split the ABCD target samples into 50 random subsets, each with about 222 persons, and merged them one at a time with the reference data, followed by running admixture k=5 on each merged subset. We verified that these subsets produced stable results by examining the stability of the estimates for the reference samples. There was very little variation across runs, e.g. for the reference sample with the most variance (European, NA12342), the mean estimate was 98.3% with *SD* = 0.17% across the 50 runs. Since Admixture does not label the resulting clusters, we used 5 reference samples to index the populations so the data would be merged correctly. In no case did this produce any inconsistencies.

1.2.2. Neuropsychological Performance

The NIH Toolbox® (NIHTBX) neuropsychological battery was designed to measure a broad range of cognitive abilities. It consists of seven tasks which index attention (Flanker Inhibitory Control and Attention Task), episodic memory (Picture Sequence Memory Task), language abilities (Picture Vocabulary Task & Oral Reading Recognition Task), executive function (Dimensional Change Card Sort Task & Flanker Inhibitory Control and Attention Task), processing speed (Pattern Comparison Processing Speed Task), and working memory (List Sorting Working Memory Task) (Akshoomoff et al., 2014; Weintraub et al., 2013). NIHTBX was normed for samples between ages 3 and 85; tasks correlate highly with comparable ability assessmnents (Weintraub et al., 2013). Moreover, this battery has been shown to be measurement invariant across American ethnic groups (Lasker, Pesta, Fuerst, & Kirkegaard, 2019). Age-corrected composite scores, based on the seven tasks, were provided by ABCD. We regressed out sex from these age-corrected composite scores. The residuals were then standardized.

1.2.3. Self-identified Race and Ethnicity

Self-identified race was based on parental responses to 18 questions asking about the child's race ("What race do you consider the child to be? Please check all that apply"). From these questions, six broad racial categories were created: European ("White"), African ("Black/African American"), Native American ("American/Native American" and "Alaska Native"), South Asian ("Asian Indian"), East Asian ("Chinese," "Filipino," "Japanese," "Korean," and "Vietnamese", "Other Asian,"), and Other ("Other race," "Refused to answer," "Don't know"). The Other Asian group (*N* = 66) was classified as "East Asian" because the Asian ancestry component was predominantly East (44%;) not South (7%) Asian; the remaining ancestry was predominantly European (40%). The Pacific Islander groups ("Native Hawaiian," "Guamanian," "Samoan," and "Other Pacific Islander") were excluded as we did not have a corresponding admixture component. Self-identified ethnicity was based on parental responses to 1 question asking about Latin American ethnicity ("Do you consider yourself Hispanic/Latino/Latina?"). From this we created an additional ethnic category.

Descriptive statistics for the SIRE groups are shown in Table S1. Statistics are reported for single ethnic categories i.e., individuals reported as being only White, Black, East Asian, Native American, or Other, with no combinations (e.g., Hispanic & White), Hispanics, the seven top double combinations (i.e., Hispanic & White, Hispanic & Black, Hispanic & Other, non-Hispanic Black & White, non-Hispanic East Asian & White, non-Hispanic Native American & White, and non-Hispanic South Asian & White) and finally all other remaining groups combined.

Table S1: Descriptive Statistics for the SIRE Groups

	N	Age *M*	Eur. %	Afr. %	E.Asian%	S.Asian %	Amer. %	US Born *N*	Neuropsy.	SES
Total Sample	9972	9.91 (0.63)	0.74 (0.31)	0.16 (0.29)	0.03 (0.12)	0.01 (0.06)	0.06 (0.13)	9703	0.00 (1.00)	0.00 (1.00)
NH White Only	5533	9.93 (0.63)	0.97 (0.05)	0.01 (0.02)	0.00 (0.02)	0.01 (0.02)	0.01 (0.03)	5459	0.25 (0.92)	0.40 (0.75)
NH Black Only	1434	9.91 (0.61)	0.18 (0.11)	0.80 (0.11)	0.00 (0.02)	0.00 (0.01)	0.01 (0.02)	1402	-0.77 (0.87)	-1.00 (0.95)
NH East Asian Only	107	10.02 (0.62)	0.14 (0.18)	0.01 (0.05)	0.82 (0.23)	0.02 (0.10)	0.01 (0.02)	88	0.57 (1.02)	0.60 (0.75)
NH South Asian Only	43	10.03 (0.68)	0.24 (0.13)	0.00 (0.00)	0.03 (0.07)	0.73 (0.14)	0.01 (0.01)	35	0.45 (1.02)	0.88 (0.46)
NH Native American Only	31	9.70 (0.60)	0.71 (0.30)	0.11 (0.26)	0.01 (0.02)	0.01 (0.01)	0.15 (0.19)	31	-0.42 (0.79)	-0.81 (0.72)
NH Other Only	97	9.96 (0.61)	0.55 (0.30)	0.28 (0.31)	0.06 (0.19)	0.04 (0.11)	0.07 (0.15)	87	-0.22 (1.12)	-0.53 (1.11)
Any Hispanic	1869	9.88 (0.63)	0.60 (0.20)	0.10 (0.14)	0.02 (0.06)	0.01 (0.02)	0.27 (0.18)	1755	-0.23 (0.96)	-0.41 (0.93)
White Hispanic	1171	9.89 (0.64)	0.67 (0.18)	0.06 (0.06)	0.01 (0.02)	0.01 (0.01)	0.26 (0.17)	1097	-0.19 (0.96)	-0.29 (0.91)
Black Hispanic	84	9.80 (0.64)	0.35 (0.13)	0.53 (0.17)	0.00 (0.01)	0.00 (0.01)	0.11 (0.09)	79	-0.34 (0.93)	-0.64 (0.98)
Other Hispanic	411	9.89 (0.63)	0.49 (0.15)	0.09 (0.11)	0.02 (0.03)	0.00 (0.01)	0.40 (0.17)	383	-0.45 (0.92)	-0.75 (0.88)
NH Black & White	302	9.88 (0.62)	0.58 (0.12)	0.41 (0.12)	0.00 (0.01)	0.00 (0.01)	0.01 (0.02)	301	-0.13 (0.95)	-0.45 (1.06)
NH East Asian & White	249	9.99 (0.64)	0.56 (0.12)	0.01 (0.01)	0.41 (0.14)	0.02 (0.04)	0.01 (0.01)	244	0.58 (0.98)	0.66 (0.67)
NH Native American & White	131	9.78 (0.60)	0.90 (0.10)	0.01 (0.03)	0.01 (0.02)	0.01 (0.01)	0.07 (0.09)	131	0.01 (0.86)	-0.09 (0.91)
NH South Asian & White	40	9.79 (0.54)	0.63 (0.11)	0.00 (0.00)	0.02 (0.08)	0.34 (0.12)	0.01 (0.01)	39	0.83 (0.84)	0.78 (0.77)
Any_Other NH Combination	136	9.87 (0.62)	0.37 (0.21)	0.46 (0.25)	0.13 (0.20)	0.02 (0.07)	0.02 (0.05)	131	-0.27 (1.06)	-0.63 (1.03)

Note: Euro.% = European ancestry percentage, Afr.% = African ancestry percentage, E.Asian% = East Asian Ancestry percentage, S.Asian% = South Asian Ancestry percentage, Neuropsy. = Neuropsychiatric performance, SES = general socioeconomic component score. NH = non-Hispanic. Hispanic & White, Hispanic & Black, and Hispanic & Other are subsets of the Any Hispanic category.

The racial and ethnic variables were then recoded to create interval categories for which individuals are assigned a percentage of each SIRE category based on the number of responses chosen (Liebler & Halpern-Manners, 2008; Kirkegaard et al., 2019). By this coding, if someone was marked as White and Hispanic, they were assigned scores of .5 for white and .5 for Hispanic and 0 for the other 5 categories. The correlations between these interval scores and genetic ancestry components are shown in Table S2 below (*N* = 9972). These associations are similar to those found by others (for example: Guo et al. 2014); self-identified race generally corresponds with genetic ancestry.

Table S2: Correlations between Interval Coded SIRE and Genetic Ancestry

	European ancestry	African ancestry	East Asian ancestry	South Asian ancestry	Amerindian ancestry
White_SIRE	**0.91**	-0.74	-0.21	-0.05	-0.31
Black_SIRE	-0.76	**0.96**	-0.08	-0.09	-0.18
East_Asian_SIRE	-0.24	-0.08	**0.92**	0.02	-0.06
South_Asian SIRE	-0.11	-0.04	0.01	**0.87**	-0.03
Native_SIRE	-0.02	0.00	-0.01	-0.03	0.07
Hispanic_SIRE	-0.21	-0.10	-0.04	-0.06	**0.77**
Other_SIRE	-0.16	-0.01	0.00	0.00	**0.38**

1.2.4. Region of Birth (US Born)

Region of birth was based on the parental response to the question, “In which country was the child born?”. The response “United States” was recoded as 1 and all other responses were recoded as 0.

1.2.5. Socioeconomic Status

Socioeconomic status was based on seven indicators: financial adversity, area deprivation index, neighborhood safety protocol, parental education, parental income, parental marital status, and parental employment status. These are detailed below:

1.2.5.1. Financial Adversity (Reverse Coded).

Parents answered a seven item Financial Adversity Questionnaire (PRFQ). They were asked: “In the past 12 months, has there been a time when you and your immediate family experienced any of the following:

(1) “Needed food but could not afford to buy it or could not afford to go out to get it?”,
(2) “Were without telephone service because you could not afford it?”
(3) “Did not pay the full amount of the rent or mortgage because you could not afford it?”,
(4) “Were evicted from your home for not paying the rent or mortgage?”,

(5) "Had services turned off by the gas or electric company, or the oil company would not deliver oil because payments were not made?",
(6) "Had someone who needed to see a doctor or go to the hospital but did not go because you could not afford it?", and
(7) "Had someone who needed a dentist but could not go because you could not afford it?"
For each of the seven items they answered "yes" (1) or "no" (0). We summed responses. Thus the maximum was 7 and the minimum was 0.

This variable was reverse coded, so that higher scores indicated less financial adversity, and then standardized.

1.2.5.2. Area Deprivation Index (ADI) (Reverse Coded)

Parents completed a residential history questionnaire. They provided the residential addresses and the number of full years they lived at each residence. For each address an Area Deprivation Index (ADI) was computed by ABCD and the national percentile of the area's socioeconomic status was given. ADI was based on the following variables: "Percentage of occupied housing units without complete plumbing (log)"; "Percentage of occupied housing units without a telephone"; "Percentage of occupied housing units without a motor vehicle"; "Percentage of single"; "Percentage of population below 138% of the poverty threshold"; "Percentage of families below the poverty level"; "Percentage of civilian labor force population aged >=16 y unemployed (unemployment rate)"; "Percentage of occupied housing units with >1 person per room (crowding)"; "Percentage of owner"; "Median monthly mortgage"; "Median gross rent"; "Median home value"; "Income disparity defined by Singh as the log of 100 x ratio of the number of households with <10000 annual income to the number of households with >50000 annual income"; "Median family income"; "Percentage of population aged >=25 y with at least a high school diploma"; "Percentage of population aged >=25 y with <9 y of education"

Scores were provided in terms of national percentiles. We used scores for the most recent residence (variable: reshist_addr1_adi_perc). The resultant values were reverse coded to make higher values indicate better neighborhoods, and then standardized.

1.2.5.3. Neighborhood Safety Protocol

Parents were asked three Likert scale (1 = strongly disagree; 5 = strongly agree) questions about neighborhood safety: "I feel safe walking in my neighborhood, day or night," "Violence is not a problem in my neighborhood," and "My neighborhood is safe from crime." We used the precomputed summary scores for which the three scores were summed and then divided them by three.

1.2.5.4. Education

Parents were asked, "What is the highest grade or level of school you have completed or the highest degree you have received." To create an interval variable, we recoded parental education as 0 to 18: Never attended/Kindergarten only = 0, 1st grade = 1, 2nd grade = 2, 3rd grade = 3, 4th grade = 4, 5th grade = 5, 6th grade = 6, 7th grade = 7, 8th grade = 8, 9th grade = 9, 10th grade = 10, 11th grade = 11, 12th grade = 12, High school graduate =12, GED or equivalent Diploma General =12, Associate degree: Occupational Program =14, Associate degree: Academic Program = 14, Bachelor's degree = 16, Master's degree = 18, Professional school = 18, Doctoral degree = 18. We standardized the scores for each educational scores for both parents and standardized the average scores.

1.2.5.5. Income

Family was an interval variable which reflected the parents' total combined family income in the past 12 months. The variable was recoded as follows: 1.00 = less than \$5000 (recode: 4,500); 2.00 = \$5000 to 11,999 (recode: 5,000); 3.00 = \$12,000 to 15,999 (recode: 12,000); 4.00 = \$16,000 to 24,999 (recode: 16,000); 5.00 = \$25,000 to 34,999 (recode: 25,000); 6.00 = \$35,000 to 49,999 (recode: 35,000); 7.00 = \$50,000 to 74,999 (recode: 50,000); 8.00 = \$75,000, to 99,999 (recode: 75,000); 9.00 = \$100,000 to 199,999 (recode: 100,000); 10.00 = \$200,000 and greater (recode: 200,000).

1.2.5.6. Marital Status

Parental marital status was coded as 1 if married and 0 for any other arrangement.

1.2.5.7. Employment Status

Parental employment was coded as 1 if at least one parent was working now either full or part time and 0 for all other cases.

1.2.5.8. General SES

Missing data for the seven economic indicators were imputed using the mice package (df, m=5, maxit = 50, method = 'pmm', seed = 500). Descriptive statistics for the imputed SES indicators are provided in Table S3, while the correlation matrix for the imputed variables ($N = 9972$), along with neuropsychiatric performance, is shown in Table S4.

Table S3: Descriptive Statistics &Correlation Matrix for the SES indicators ($N = 9972$ in all cases)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Financial Adversity	0.00	1.00								
2. ADI	0.00	1.00	.21**							
3. Neighbourhood Safety Protocol	0.00	1.00	.24**	.25**						
4. Education	0.00	1.00	.28**	.28**	.28**					
5. Income	0.00	1.00	.34**	.34**	.31**	.57**				
6. Marital Status	0.68	0.47	.27**	.23**	.23**	.33**	.46**			
7. Employment Status	0.91	0.28	.16**	.13**	.15**	.26**	.27**	.28**		
8. General SES	0.00	1.00	.56**	.54**	.55**	.73**	.80**	.66**	.48**	
9. Neuropsy.	0.00	1.00	.20**	.21**	.18**	.40**	.35**	.27**	.18**	.43**

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

We then submitted the seven SES indicators to Principal Component Analysis (PCA). For this, we used the R package PCAmixdata, which handles mixed categorical and continuous data (Chavent, Kuentz-Simonet, & Saracco, 2014). The first unrotated component explained 39% of the variance. The PCA_1 loadings for the seven SES indicators were as follows: financial adversity (.317), area deprivation index (.296), neighborhood safety protocol (.298), parental education (.529), parental income (.641), parental marital status (.431), and parental employment status (0.230). The vector of PCA_1 loadings correlated with the vector of SES indicator effects on Neuropsychiatric Performance at $r = .91$ ($N = 7$). Similarly, the vector of PCA_1 loadings strongly correlated with the vectors of genetic ancestry, with absolute values of $r = .62$ to $r = .95$ ($N = 7$). Table S4 shows the correlations between SES indicators and genetic ancestry for all five ancestry components. These results indicate that the better measures of general SES have stronger correlations with both Neuropsychiatric Performance and genetic ancestry.

Table S4: General SES-loadings and the Correlations between SES Indicators and Genetic Ancestry within the Full Sample

	General SES Loadings	*r*Europe	*r*African	*r*Amerindian	*r*East Asian	*r*South Asian
Financial Adversity	0.317	0.24	-0.26	-0.07	0.05	0.05
ADI	0.296	0.26	-0.32	0.00	0.06	0.04
Neighbourhood Safety Protocol	0.298	0.30	-0.28	-0.12	0.02	0.04
Education	0.529	0.38	-0.29	-0.39	0.10	0.12
Income	0.641	0.38	-0.36	-0.23	0.09	0.11
Marital Status	0.431	0.37	-0.40	-0.09	0.07	0.07
Employment Status	0.230	0.21	-0.20	-0.08	0.02	0.03
***r*(SES loadings x ancestry correlations)**		**0.89**	**-0.62**	**-0.73**	**0.87**	**0.95**

Note: Of relevance is the magnitude, not direction, of the correlations between the vector of SES loadings and the vectors of ancestry correlations.

PCA_1 scores correlated with Neuropsychological Performance at $r = .43$ in the full sample. Among non-Hispanic Whites, non-Hispanic Blacks, and Hispanics, the correlations between PCA_1 and Neuropsychological Performance was $r = .25$ ($N = 5533$), $r = .31$ ($N = 1434$), and $r = .31$ ($N = 1869$), respectively. These magnitudes of child-parental SES correlations are consistent with those previously reported (Flores-Mendoza, Ardila, Gallegos, & Reategui-Colareta, 2021; Sirin, 2005). The congruence coefficients for the SES component loadings were greater or equal to $r = .97$ for the largest three SIRE groups (non-Hispanic Whites, non-Hispanic Blacks, and Hispanics), indicating identical structures across groups.

1.3. Methods

A series of regression models were run with NIHTBX as the dependent variable. The NIHTBX and socioeconomic variables were standardized (based on the subsample of 9972 retained). The ancestry variables were left unstandardized, thus the coefficients from ancestries can be interpreted as a change in 100% ancestry over a change in one standardized unit of NIHTBX scores. European ancestry and White SIRE were selected as reference values and thus not included as independents. For the regression analyses, following the recommendations of Heeringa and Berglund (2021), we used a three-level (site, family, individual) multi-level mixed effects model. This model was applied to the pooled twin and regular ABCD baseline sample. This specification approximates the ABCD Data Exploration and Analysis Portal (DEAP) specification (Heeringa and Berglund, 2021).

References

Akshoomoff, N., Newman, E., Thompson, W. K., McCabe, C., Bloss, C. S., Chang, L., ... & Jernigan, T. L. (2014). The NIH Toolbox Cognition Battery: results from a large normative developmental sample (PING). *Neuropsychology*, 28(1), 1.

Chavent, M., Kuentz-Simonet, V., & Saracco, J. (2014). Multivariate analysis of mixed data: the PCAmixdata R package. arXiv. arXiv preprint arXiv:1411.4911.

Flores-Mendoza, C., Ardila, R., Gallegos, M., & Reategui-Colareta, N. (2021) General Intelligence and Socioeconomic Status as Strong Predictors of Student Performance in Latin American Schools: Evidence From PISA Items. *Front. Educ*. 6:632289. doi: 10.3389/feduc.2021.632289

Guo, G., Fu, Y., Lee, H., Cai, T., Harris, K. M., & Li, Y. (2014). Genetic bio-ancestry and social construction of racial classification in social surveys in the contemporary United States. *Demography*, 51(1), 141-172.

Heeringa, S. G., & Berglund, P. A. (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. *BioRxiv*.

CHAPTER 9

GENETIC ANCESTRY AND GENERAL COGNITIVE ABILITY IN A SAMPLE OF AMERICAN YOUTHS

JOHN G.R. FUERST, MENG HU, GREGORY CONNOR

Abstract

Black and Hispanic children in the United States have lower mean cognitive test scores than White children. The reasons for this are contested. The test score gap may be caused by socio-cultural factors, but the high heritability of *g* suggests that genetic variance might play a role. Differences between self-identified race or ethnicity (SIRE) groups could be the product of ancestral genetic differences. This genetic hypothesis predicts that genetic ancestry will predict *g* within these admixed groups. To investigate this hypothesis, we performed admixture-regression analyses with data from the Adolescent Brain Cognitive Development Cohort. Consistent with predictions from the genetic hypothesis, African and Amerindian ancestry were both found to be negatively associated with *g*. The association was robust to controls for multiple cultural, socioeconomic, and phenotypic factors. In the models with all controls the effects were as follows: (a) Blacks, African ancestry: b = -0.89, *N* = 1690; (b) Hispanics, African ancestry: b = -0.58, Amerindian ancestry: b = -0.86, *N* = 2021), and (c) a largely African-European mixed Other group, African ancestry: b = -1.08, *N* = 748). These coefficients indicate how many standard deviations *g* is predicted to change when an individual's African or Amerindian ancestry proportion changes from 0% to 100%. Genetic ancestry statistically explained the self-identified race and ethnicity (SIRE) differences found in the full sample. Lastly, within all samples, the relation between genetic ancestry and *g* was partially accounted for by cognitive ability and educational polygenic scores (eduPGS). These eduPGS were found to be significantly predictive of *g* within all SIRE groups, even when controlling for ancestry. The results are supportive of the genetic model.

Keywords: ancestry, admixture, education, intelligence, ABCD cohort

1. Introduction

There are substantial differences in mean cognitive test scores between self-identified racial and ethnic (SIRE) groups such as Whites, Hispanics, Blacks, and Asians in the United States (Murray, 2021; Roth et al., 2017). These differences are not attributable to psychometric bias, since cognitive test batteries typically exhibit measurement invariance across American SIRE groups (Scheiber, 2016a,b; Warne, 2020a). As such, they represent real differences in latent cognitive ability. Although cognitive ability researchers disagree as to the causes of these differences (Rindermann, Becker & Coyle, 2020), a number of explanatory factors have been proposed. Most of these factors appeal to either cultural differences between groups or differing social and economic circumstances such as differences in poverty levels. Many researchers studying cognitive ability attribute some part of the differences to genetics (Rindermann, Becker & Coyle, 2020).

Analyses of both nationally representative samples and cognitive battery standardization data (Magnuson & Duncan, 2006; Weiss & Saklofske, 2020) indicate that socioeconomic status (SES) can statistically explain a substantial percentage of test score variance across SIRE groups. However, it is not clear to what extent SES captures predominantly environmental or genetic causes, since SES is related to both genetic and environmental differences within groups (Belsky et al., 2018; Krapohl & Plomin, 2016; Rowe, Vesterdal & Rodgers, 1998). Regardless, the data suggest that even after fully accounting for average SES differences, there remains unexplained variance in cognitive test scores between groups.

The genetic hypothesis is that differences in genes inherited from ancestors play a significant role in causing the SIRE related variation in cognitive test scores. Twin studies show non-trivial heritability for individual differences in cognitive ability within ethnic groups (Pesta et al., 2020). SIRE groups differ in European ancestry-based polygenic scores, and these scores are predictive of cognitive ability within ethnic groups (Lasker et al., 2019). Taken together, these results suggest that variation in cognitive ability among SIRE groups may be due in part to allele frequency differences at trait-associated gene loci.

Most studies that investigate the source of group differences categorize individuals by SIRE. However, SIRE reflects both genetic and sociocultural factors. This makes interpretation of simple SIRE-based results ambiguous. Admixture regression analyses, by including both SIRE and admixture variables simultaneously, can quantify the association between genetic ancestry and phenotype in admixed populations (Halder et al., 2015). These analyses test the extent to which the genetic differences between SIRE groups are responsible for observed trait differences.

If continental populations differ polygenically in trait-related phenotypes, these continental differences are expected to contribute to individual differences in admixed populations. In aggregate, alleles that vary between ancestral populations will be associated with phenotype in admixed populations. Individuals who have a larger proportion of their global ancestry from the ancestral group with higher average frequencies of trait-enhancing alleles are likely to also have higher polygenic scores and higher average values in the phenotype. As such, associations between genetic ancestry and phenotype in admixed populations would suggest that phenotypic differences between continental populations have a genetic basis (Halder et al., 2015). This admixture regression methodology can be extended by including polygenic scores (PGS) to see if PGS mediate the association between *g* and ancestry.

In this way, admixture-regression analysis allows researchers to separate the genetic element of SIRE from its cultural, behavioral, and psychosocial aspects, which may alternatively be responsible for the observed phenotypic differences. As Fang et al. (2019, p. 764) note: SIRE "acts as a surrogate to an array of social, cultural, behavioral, and environmental variables" and so "stratifying on SIRE has the potential benefits of reducing heterogeneity of these non-genetic variables and decoupling the correlation between genetic and non-genetic factors."

These analyses require a substantial degree of admixture in populations, and are more robust when that admixture has taken place in the course of the last seven to ten generations (Halder et al., 2015). In the USA, African and Hispanic Americans meet these requirements. They exhibit a wide range of European, African, and/or Amerindian ancestry due to admixture over the course of several generations.

Thus we apply admixture analysis to examine if SIRE differences in general cognitive ability (*g)* can be accounted for by genetic variation related to continental ancestry. We hypothesize that *g* will be lower in African and Hispanic American samples relative to European American samples. These differences will be associated with African and Amerindian genetic ancestry within the SIRE groups. Additionally, we hypothesize that the association between genetic ancestry and *g* will be robust to controls for possible socio-environmental confounds and that genetic ancestry will also statistically account for the SIRE differences found in the full sample. Finally, we hypothesize that current polygenic scores for cognitive ability and education (eduPGS) will both predict individual differences within SIRE groups and explain a portion of the effect of ancestry on *g*.

2. Methods

2.1. Dataset

The Adolescent Brain Cognitive Development Study (ABCD) is a collaborative longitudinal project involving 21 sites across the USA. ABCD is the largest, longitudinal study of brain development and child health ever conducted in the USA. It was created to research the psychological and neurobiological bases of human development. At baseline, around 11,000 children aged 9-10 years were sampled, mostly from public and private elementary schools. A probabilistic sampling strategy was used, with the goal of creating a broadly representative sample of US children in that age range. Children with severe neurological, psychiatric, or medical conditions were excluded. Children were also excluded if they were not fluent in English or if their parents were not fluent in either English or Spanish. Parents provided informed consent. For this study, we utilized the baseline ABCD 3.01 data. We excluded individuals missing either cognitive or admixture scores. We also excluded any individual identified as being either Asian or Pacific Islander in order to focus on groups who were primarily of African, European, and Amerindian ancestry. This left 10,370 children.

For the admixture-regression analyses, SIRE groups were delineated using the ABCD race_ethnicity variable. This was a summary variable computed from 18 separate multiple choice questions asking about the child's race ("What race do you consider the child to be? Please check all that apply") and one question asking if the child is of Hispanic ethnicity ("Do you consider yourself Hispanic/Latino/Latina?"). Children were classified into one of five mutually exclusive groups: non-Hispanic White (*White*), non-Hispanic Black (*Black*), Hispanic of any race (*Hispanic*), non-Hispanic Asian (*Asian*) or any other (*Other*). The Other category included any non-Hispanic children who were reported to be two or more racial groups. Because we dropped the Asian category, we were left with four mutually exclusive SIRE groups.

2.2. Variables for admixture regression analyses

The following variables were used for the admixture regression analyses:

2.2.1. *g* scores

ABCD baseline data contain the following cognitive subtests, the first seven of which are from the NIH Toolbox® cognitive battery: Picture Vocabulary, Flanker, List Sorting, Card Sorting, Pattern Comparison, Picture Sequence Memory, Oral Reading Recognition, Wechsler Intelligence Scale for Children's Matrix Reasoning, The Little Man Test (efficiency score), The Rey Auditory Verbal Learning Test (RAVLT), immediate recall, and RAVLT delayed recall. For details about these measures, see Thompson et al. (2019).

We conducted multi-group confirmatory factor analysis (MGCFA) on these subtests, as detailed in Supplementary File 1. Briefly, we first checked whether outliers and missing data had any impact, and whether our results remained strong after correction. We then conducted exploratory factor analysis and multi-group confirmatory factor analysis on the aforementioned set of subtests as a check for bias. After adjustment for age, we did not find any non-linear effects of age. Adjustment for sex did not reveal any evidence of meaningful differences in fit between the competing models, the *g*-model and the correlated factors model. We find that a three broad factor model (memory, complex cognition, and executive function) with *g* at the apex fits the data well. Moreover, strict measurement invariance holds between SIRE groups. The best fitting model (M6A, Table S2 of Supplementary File 1; CFI = .954, RMSEA = .044) was one in which *g* alone explains SIRE group differences. We output the *g*-factor scores from this model for use in the analyses. These score magnitudes are approximately the same as those derived from exploratory factor analysis.

2.2.2. Socioeconomic status (SES)

We identified seven indicators of SES: financial adversity, area deprivation index, neighborhood safety protocol, parental education, parental income, parental marital status, and parental employment status. These are detailed in Supplementary File 2. We submitted the seven SES indicators to Principal Components Analysis (PCA). We used the R package PCAmixdata, which handles mixed categorical and continuous data (Chavent et al., 2014). The first unrotated component explained 42% of the variance. The PCA_1 loadings for the seven SES indicators were as follows: financial adversity (.31), area deprivation index (.49), neighborhood safety protocol (.31), parental education (.53), parental income (.66), parental marital status (.42), and parental employment status (0.21). More details and the correlation matrix for the SES indicators is provided in Supplemental File 2.

2.2.3. Child US-born

Parents were asked about the country of the child's birth. We recoded this variable as 1 for "United States" and 0 for all other responses.

2.2.4. Immigrant family

Parents were asked if anyone in the child's family, including maternal or paternal grandparents, was born outside of the United States. This variable was coded as 1 for "Yes" and 0 for all other responses.

2.2.5. Nationality (Puerto Rican, Mexican, and Cuban)

If a child was reported to be Hispanic, parents were additionally asked about the specific Latin American nation of origin ("Please choose the group that best represents the child's Hispanic origin or ancestry"). Seventy percent of the Hispanic children were reported as being either Mexican, Mexican American, or Chicano ($N = 1028$), Puerto Rican ($N = 210$), or Cuban or Cuban American ($N = 174$). Dummy variables were created for these three nationality groups, with "1" indicating "Yes" and "0" indicating "No".

2.2.6. Frac_SIRE

Four dummy SIRE variables (Black, White, Native American, and Not Otherwise Classified (NOC) were computed from the 18 questions asking about the child's specific race. The NOC SIRE group included those who were marked as: "Other Race," "Refused to answer," or "Don't Know." These were then recoded into interval variables in which individuals are assigned a SIRE fraction ranging from 0 to 1 (Liebler & Halpern-Manners, 2008). These were calculated as the value selected for each of the four groups (0 or 1) over the total number of responses (0 to 4) chosen. For example, someone marked as only Black and White would be assigned scores of (Black: ½; White: ½; Native_American: 0; NOC = 0). This SIRE coding was used as it was previously found to be the most predictive in models which also included genetic ancestry (Kirkegaard et al., 2019).

2.2.7. Hispanic

For the admixture-regression analysis conducted on the full sample, we additionally included a dummy variable for Hispanic ethnicity. This was coded as "1" for "Hispanic" and "0" for not Hispanic. As the subsamples were either Hispanic or non-Hispanic, this variable was not used in the subsample analyses.

2.2.8. Ethnic attachment

Parents were given the Multigroup Ethnic Identity Measure-Revised (MEIM-R) Survey. In this they were asked six Likert-scaled (1 = strongly agree; 5 = strongly disagree) questions regarding their ethnic group: "I have spent time trying to find out more about my ethnic group, such as its history, traditions, and customs", "I have a strong sense of belonging to my own ethnic group," "I understand pretty well what my ethnic group membership means to me," "I have often done things that will help me understand my ethnic background better,", "I have often talked to other people in order to learn more about my ethnic group,", "I feel a strong attachment towards my own ethnic group." ABCD computed MEIM-R summary scores, which we standardized. We treat this as a measure of family ethnic-related culture. We only included this variable in the subsample analyses. In these the members belonged to the same broad ethnic group (e.g., Black or Hispanic).

2.2.9. State racism

ABCD calculated state-level indicators of both racism and immigrant bias. These were based on both implicit bias measures and state-level structural variables. The two indicators correlated at $r = .41$ ($N = 9386$). We standardized both measures ($M = 0$, $SD = 1$) and then averaged them and standardized the resulting average.

2.2.10. Discrimination factor

In Year 1 follow-up, the children were asked 6 questions regarding perceived ethnic, racial, national, or color based discrimination. The questions were as follows: "In the past 12 months, have you felt discriminated against: because of your race, ethnicity, or color?", "In the past 12 months, have you felt discriminated against: because you are (or your family is) from another country?", "How often do the following people treat you unfairly or negatively because of your ethnic background?"

(Teachers? Other adults outside school? Other students?), "I feel that others behave in an unfair or negative way toward my ethnic group." We imputed missing data using the mice package (df, m = 5, maxit = 50, method = 'pmm', seed = 500). We used the mirt package in R to perform factor analysis on the six questions. We then standardized and saved the factor scores.

2.2.11. Skin_color, P_Brown_Eye, P_Intermediate_Eye, P_Blue_Eye, P_Black_Hair, P_Brown Hair, P_Red_or_Blond_Hair

Conley and Fletcher (2017) have suggested that phenotypic-based discrimination might mediate the association between cognitive ability and genetic ancestry. This is called the colorism model (Hu et al., 2019). It can be tested by including indices of race-related phenotype into the regression models to see if these capture the association between ancestry and cognitive ability. As such, we include measures of eye, hair, and skin color. Skin, Hair, Eye color were calculated based on the publicly available, "Hirisplex Eye, Hair and Skin Colour DNA Phenotyping Webtool." This tool and score calculations have been detailed by Lasker et al. (2019). We additionally combined (summed) the red and blond hair probabilities. Skin color was scaled as detailed in Lasker et al. (2019), with higher scores representing darker color, and then standardized. The eye and hair color variables represent the percent in the full sample with the specific color and were left unstandardized to retain interpretability.

2.2.12. Admixture estimates

Imputing and genotyping was done by the ABCD Research Consortium using Illumina XX. 516,598 variants survived the quality control. Before global admixture estimation, we applied quality control using PLINK 1.9. We used only directly genotyped, bi-allelic, autosomal SNP variants (494,433 before, 493,196 after lifting). We pruned variants for linkage disequilibrium at the 0.1 R^2 level using PLINK 1.9 (--indep-pairwise 10000 100 0.1). This variant filtering was done in the reference population dataset to reduce bias from sample non-representativeness. 99,642 variants were left after pruning. We merged the target samples from ABCD with reference population data for the populations of interest. A k=5 solution with European, Amerindian, African, East Asian and South Asian components provides the most comprehensive but parsimonious model of the US population, capturing all the predominant ancestral backgrounds in the US population. We merged our sample with relevant samples from 1000 Genomes and from the HGDP to perform the cluster analysis and identify these k=5 components. The following populations from 1000 Genomes and from the HGDP reference populations were excluded: Adygei, Balochi, Bedouin, Bougainville, Brahui, Burusho, Druze, Hazara, Makrani, Mozabite, Palestinian, Papuan, San, Sindhi, Uygur, Yakut. We excluded these populations because they were overly admixed or because the individuals in the ABCD sample lacked significant portions of these ancestries (e.g., Melanesians and San). We split the ABCD target samples into 50 random subsets (222 persons each) and merged them sequentially with the reference data. Admixture at k = 5 was run on each of the 50 merged subsets. This repeated subsetting was done to avoid skewing the admixture algorithm to European ancestry which was predominant in the ABCD sample.

2.2.13. First 20 Principal components

For the analysis of PGS predictivity within SIRE groups we controlled for the first 20 ancestry principal components to take into account population structure related effects. These components were generated by PLINK v1.90b6.8 when computing polygenic scores.

2.2.14. eduPGS

For polygenic scores (PGS), we scored the genomes using PLINK v1.90b6.8. For background, a polygenic score (PGS) "is an estimate of an individual's genetic liability to a trait or disease, calculated according to their genotype profile and relevant genome-wide association study (GWAS) data" (Choi, Mak & O'Reilly, 2020). We used the genome-wide association study (GWAS) results from Lee et al. (2018). Specifically, we used the multi-trait analysis of genome-wide association study (MTAG) eduPGS SNPs ($N = 8{,}898$ variants in this sample) to compute eduPGS. The MTAG eduPGS were computed using MTAG, a method for analyzing statistics from genome-wide association studies (GWAS) on different but genetically correlated traits (e.g., education and intelligence). These scores were based on cognitive ability ($N = 257{,}841$), hardest math class taken ($N = 430{,}445$), and mathematical ability ($N = 564{,}698$) (Lee et al., 2018). We use these PGS because previous research has shown them to have trans-ethnic predictive validity in European, Hispanic, and African American populations (Fuerst, Kirkegaard & Piffer, 2021; Lasker et al., 2019). Moreover, common forms of bias were found not to account for the ancestry-related eduPGS differences (Fuerst et al., 2021). Thus, we can say that this PGS *plausibly* captures genetic effects between ancestry groups.

2.2.15. The NIH Toolbox® (NIHTBX) neuropsychological battery

For one validation analysis of the eduPGS which included Asians, we used the NIHTBX summary scores. This was because we did not run MGCFA on the small Asian samples and so did not have *g* scores for these groups. This battery has been shown to be measurement invariant across American Black, Hispanic, and White SIRE groups (Lasker et al., 2019). The effects of age and sex were controlled for. We standardized the residuals.

2.3. Methods (Analyses)

We first present the descriptive statistics for the sample and the subsamples. We then explore the bivariate relation between European admixture and *g*-scores. We include both linear regression lines and loess lines in the regression plots (based on the gg_scatter package in R). These analyses are descriptive and do not take into account the complex structure of the data.

After, we run a series of within-SIRE (Black, Hispanic, and Other) admixture-regression analyses to control for potential environmental confounds. For these analyses, we set European ancestry as a reference value with a value of zero. Following Heeringa and Berglund's (2020) recommendations, we use a multi-level mixed effects three-level (site, family, individual) model. In this model, recruitment site and family common factors are treated as random effects (i.e., as random samples from a population). We further report the dense numeric matrix results for the regression models in Supplementary File 3.

The pooled data with both the regular ABCD baseline sample and the pooled twin samples were used. As Heeringa and Berglund (2020) note, the specification replicates that used by the ABCD Data Exploration and Analysis Portal (DEAP). Thus, the use of this multilevel model also aids in replication. For the regression analyses on the SIRE subsamples, we ran four models. The first model includes genetic ancestry and controls for both child and family immigrant status. The second model adds a term for SIRE and ethnic attachment to capture SIRE specific cultural effects. The third model adds terms to capture possible discrimination related effects: state-level racism, child reported experiences of discrimination, and race-related phenotype. The fourth model adds our general SES variable. Geographic effects are controlled for by including study site as a random effect in the model.

For the regression analyses, general cognitive ability scores (*g*-scores) are used as the dependent variable. This variable was standardized ($M = 0.00$; $SD = 1.00$) in the full sample. As for the independent variables, both the ancestry and fractional SIRE variables were left unstandardized. This allows the unstandardized beta coefficients for these variables to be interpreted as the effect of a change in 100 percent ancestry/SIRE identity on one standardized unit of cognitive ability. The rationale for this method has been detailed elsewhere (Lasker et al., 2019). The Child_USA_Born and Immigrant_Family dummy variables were also left unstandardized to retain interpretability. The three eye color and the three hair color variables, which represent probabilities that sum to one in the full sample, are also not standardized to retain interpretability. The remaining variables — ethnic attachment, state racism, discrimination factor, skin color, and SES — are all standardized in the full sample. Thus, the unstandardized B coefficient for these variables represents the change in *g* induced by a change of one standard deviation in the independent variable.

3. Results

3.1. Descriptive statistics

The descriptive statistics for the total sample and the four SIRE subsamples are shown in Table 1. Cohen's *d* for the difference in *g* between Black and White Americans comes to 1.02 *d*. This represents a large effect by conventional standards (Cohen, 1988) and is typically sized for measured g differences across SIRE groups (Roth et al., 2017). The difference between Hispanic and White Americans is 0.38 *d*, while that between Others and White Americans is 0.37 *d*. These latter two differences represent small to medium sized effects (Cohen, 1988). The Hispanic-White difference is smaller than usually reported (e.g., Roth et al., 2017). This could be due to the exclusion of children who were not fluent in English.

In this sample, parent-identified Whites are 98% European in ancestry (1% African; 1% Amerindian). Since this group has little admixture, we relegate the within SIRE admixture-regression analyses to the supplementary file. Both the Black (82% African, 16% European, 1% Amerindian) and the Other (62% European, 32% African, 4% Amerindian) groups are African-European admixed groups. Hispanics additionally have a substantial Amerindian component (60% European, 28% Amerindian, 10% African). Figure 1 shows the distribution of ancestry by SIRE groups. These admixture percentages correspond with those typically reported in the literature (e.g., Bryc et al., 2015).

Table 1. Total sample and subsample characteristics

	Total sample	**Black**	**Hispanic**	**Other**	**White**
	M ± *SD*	*M* ± *SD*	*M* ± *SD*	*M* ± *SD*	*M* ± *SD*
Age (in Months)	119.0 ± 7.49	118.9 ± 7.28	118.6 ± 7.58	118.7 ± 7.40	119.2 ± 7.52
g	0.00 ± 1.00	-0.69 ± 1.07	-0.10 ± 0.99	-0.09 ± 1.09	0.24 ± 0.86
SES	0.00 ± 1.00	-0.98 ± 0.93	-0.36 ± 0.91	-0.40 ± 1.00	0.45 ± 0.75
frac_White_SIRE	0.73 ± 0.43	0.00 ± 0.00	0.67 ± 0.45	0.39 ± 0.25	1.00 ± 0.03
frac_Black_SIRE	0.20 ± 0.38	1.00 ± 0.04	0.07 ± 0.23	0.29 ± 0.26	0.00 ± 0.00
frac_Native_Amer._SIRE	0.02 ± 0.11	0.00 ± 0.00	0.03 ± 0.13	0.18 ± 0.28	0.00 ± 0.00
frac_Other_SIRE	0.06 ± 0.23	0.00 ± 0.04	0.23 ± 0.42	0.13 ± 0.34	0.00 ± 0.03
European_ancestry	0.75 ± 0.33	0.16 ± 0.11	0.60 ± 0.21	0.62 ± 0.25	0.98 ± 0.05
African_ancestry	0.18 ± 0.31	0.82 ± 0.11	0.10 ± 0.15	0.32 ± 0.26	0.01 ± 0.02
Amerindian_ancestry	0.06 ± 0.14	0.01 ± 0.02	0.28 ± 0.19	0.04 ± 0.09	0.01 ± 0.03
South_Asian_ancestry	0.00 ± 0.02	0.00 ± 0.01	0.01 ± 0.01	0.01 ± 0.05	0.00 ± 0.01
East_Asian_ancestry	0.01 ± 0.03	0.01 ± 0.02	0.01 ± 0.02	0.01 ± 0.07	0.00 ± 0.02
State_racism	0.00 ± 1.00	0.44 ± 0.93	-0.35 ± 0.93	0.26 ± 0.99	-0.04 ± 0.99
Discrim_fact	0.00 ± 1.00	0.47 ± 1.24	0.09 ± 1.04	0.19 ± 1.05	-0.19 ± 0.83
Ethnic_attachment	0.00 ± 1.00	0.38 ± 1.03	0.23 ± 1.01	0.05 ± 1.00	-0.19 ± 0.94
Skin_color	0.00 ± 1.00	1.32 ± 0.42	0.58 ± 0.80	0.36 ± 0.91	-0.62 ± 0.60
P_Brown_Eye	0.56 ± 0.41	0.97 ± 0.10	0.83 ± 0.28	0.73 ± 0.35	0.33 ± 0.35
P_Intermediate_Eye	0.08 ± 0.07	0.02 ± 0.03	0.06 ± 0.06	0.07 ± 0.06	0.10 ± 0.07
P_Blue_Eye	0.36 ± 0.41	0.02 ± 0.08	0.11 ± 0.25	0.20 ± 0.33	0.57 ± 0.39
P_Black_Hair	0.23 ± 0.24	0.55 ± 0.16	0.38 ± 0.23	0.28 ± 0.20	0.09 ± 0.10
P_Brown Hair	0.46 ± 0.18	0.43 ± 0.13	0.49 ± 0.16	0.54 ± 0.15	0.45 ± 0.20
P_Red_or_Blond_Hair	0.30 ± 0.29	0.02 ± 0.05	0.13 ± 0.18	0.18 ± 0.21	0.46 ± 0.26
Child_USA_Born	0.98 ± 0.15	0.98 ± 0.15	0.94 ± 0.24	0.98 ± 0.14	0.99 ± 0.12
Immigrant_family	0.28 ± 0.45	0.14 ± 0.35	0.73 ± 0.44	0.20 ± 0.40	0.18 ± 0.38
Puerto_Rican			0.10 ± 0.31		
Mexican			0.51 ± 0.50		
Cuban			0.09 ± 0.28		
eduPGS	0.00 ± 1.00	-1.33 ± 0.57	-0.21 ± 0.76	-0.36 ± 0.85	0.50 ± 0.77

Note: Nationality variables (Mexico, Cuba, & Puerto Rico) were only computed for Hispanics.

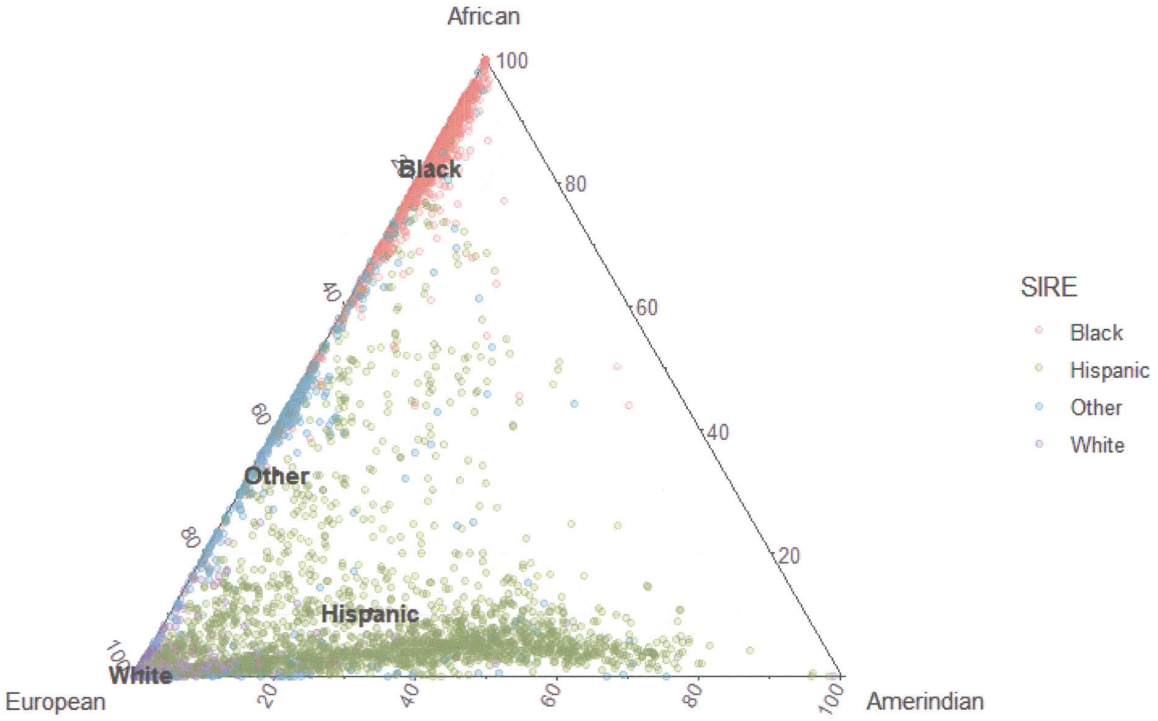


Figure 1. Admixture triangle plot for SIRE groups in the ABCD sample

3.2. Regression plots and admixture regression analyses

3.2.1. Black Americans

Figure 2 shows the regression plot for European ancestry and *g*-scores among Black children. European ancestry is significantly ($r = .10$, $N = 1690$) associated with *g* scores. The R boxplot function indicated 13 outliers. However, removing these had no effect on the bivariate correlation ($r = .10$, $N = 1677$). Additionally, the Loess regression line indicated a possible curvilinear relation with a slight uptick in scores at the lowest European ancestry decile. Further analysis showed that this was due to relatively high scores of individuals from African immigrant families ($M_{\text{African\_immigrant}} = -.28$, $N = 60$). Limiting our scope to African Americans within US-born families raises the correlation to $r = .13$ ($N = 1475$); for African Americans with 2% to 80% European admixture, this correlation is r = .11 ($N = 1635$). These results are shown in Table S4 of Supplementary file 3 along with scores by African American subgroups. The full correlation matrices are also provided in Supplementary File 3.

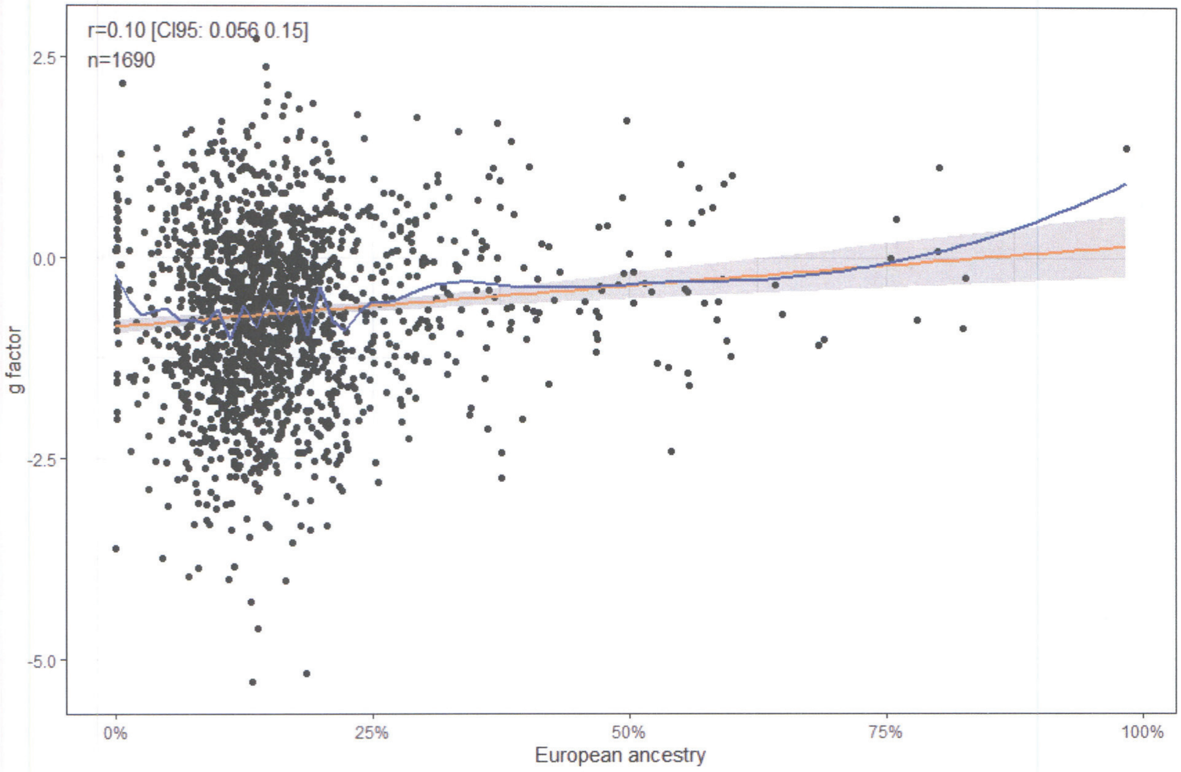


Figure 2. Regression plot of European ancestry and *g* in the Black American subsample ($N = 1690$)

We next proceed to the admixture-regression analyses. Because the Black SIRE category excludes multi-racial individuals, we do not include a term for fraction SIRE in these models. As seen in Table 2, African ancestry is strongly and significantly negatively related to cognitive ability in all four models. Amerindian ancestry is also negatively related to *g*-scores; however, owing to the low Amerindian admixture among non-Hispanic Blacks — and consequently the high standard errors — these estimates are not reliable. Adding ethnic attachment scores in Model 2 does not change the relationship with Amerindian and African ancestry. As seen in Model 3, measures of racial discrimination do not mediate the relation between *g* and ancestry. Of these variables added to Model 3, only experiences of discrimination had a significant independent effect. Finally, as seen in Model 4, while SES was significantly related to *g*, it did not substantially attenuate the association between African ancestry and *g* ($b_{\text{African ancestry}} = -1.08$ to -0.89).

Table 2. Regression results for the effect of genetic ancestry on *g* Among Black Americans ($N = 1690$). Shown are the beta coefficients (b) and *p*-values (p) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom

	Model 1		**Model 2**		**Model 3**		**Model 4**	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.00 (0.29)	0.997	0.02 (0.29)	0.934	0.69 (0.74)	0.351	0.31 (0.72)	0.669
Amerindian_ancestry	-3.80 (1.22)	**0.002**	-3.95 (1.22)	**0.001**	-3.83 (1.23)	**0.002**	-3.46 (1.20)	**0.004**
African_ancestry	-1.07 (0.26)	**<0.001**	-1.11 (0.26)	**<0.001**	-1.08 (0.29)	**<0.001**	-0.89 (0.29)	**0.002**
East_Asian_ancestry	-0.03 (1.57)	0.982	-0.07 (1.56)	0.965	0.16 (1.56)	0.920	-0.29 (1.50)	0.850
South_Asian_ancestry	-5.67 (3.42)	0.098	-5.83 (3.40)	0.087	-5.97 (3.42)	0.080	-5.46 (3.32)	0.100
Child_US_Born	0.19 (0.18)	0.287	0.16 (0.18)	0.365	0.17 (0.18)	0.354	0.32 (0.18)	0.067
Immigrant_Family	0.29 (0.08)	**<0.001**	0.27 (0.08)	**0.001**	0.27 (0.08)	**0.001**	0.15 (0.08)	0.058
Ethnic_attachment			0.11 (0.03)	**<0.001**	0.11 (0.03)	**<0.001**	0.08 (0.03)	**0.001**
State_racism					0.01 (0.05)	0.866	0.04 (0.05)	0.456
discrim_fact					-0.05 (0.02)	**0.014**	-0.04 (0.02)	**0.027**
Skin_color					0.08 (0.07)	0.290	0.11 (0.07)	0.124
P_Brown_Eye					-0.32 (0.42)	0.455	-0.33 (0.41)	0.430
P_Intermediate_Eye					-0.39 (1.76)	0.825	0.61 (1.72)	0.721
P_Black_Hair					-0.42 (0.77)	0.589	-0.12 (0.75)	0.874
P_Brown_Hair					-0.56 (0.82)	0.496	-0.26 (0.80)	0.747
SES							0.29 (0.03)	**<0.001**
Random Effects								
σ^2	0.56		0.56		0.57		0.56	
τ_{00}	$0.55_{\text{site\_id\_l:rel\_family\_id}}$ $0.03_{\text{site\_id\_l}}$		$0.53_{\text{site\_id\_l:rel\_family\_id}}$ $0.03_{\text{site\_id\_l}}$		$0.52_{\text{site\_id\_l:rel\_family\_id}}$ $0.03_{\text{site\_id\_l}}$		$0.46_{\text{site\_id\_l:rel\_family\_id}}$ $0.04_{\text{site\_id\_l}}$	
ICC	0.51		0.50		0.49		0.47	
N	$22_{\text{site\_id\_l}}$ $1436_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_l}}$ $1436_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_l}}$ $1436_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_l}}$ $1436_{\text{rel\_family\_id}}$	
Observations	1690		1690		1690		1690	
Marginal R^2 / Conditional R^2	0.022 / 0.517		0.033 / 0.515		0.037 / 0.512		0.096 / 0.520	

3.2.2. Hispanic Americans

Figure 3 shows the regression plot for European ancestry and *g* scores among Hispanic children. As seen, European ancestry is significantly ($r = .23$, $N = 2021$) associated with *g* scores. While the R boxplot function indicates that there are 23 outliers, removing these had little effect on the bivariate correlation ($r = .22$, $N = 1998$). The Loess regression line suggests a possible slight uptick in scores at the lowest European ancestry decile. However, the 95% confidence intervals of this line (not shown) overlapped with the linear regression line.

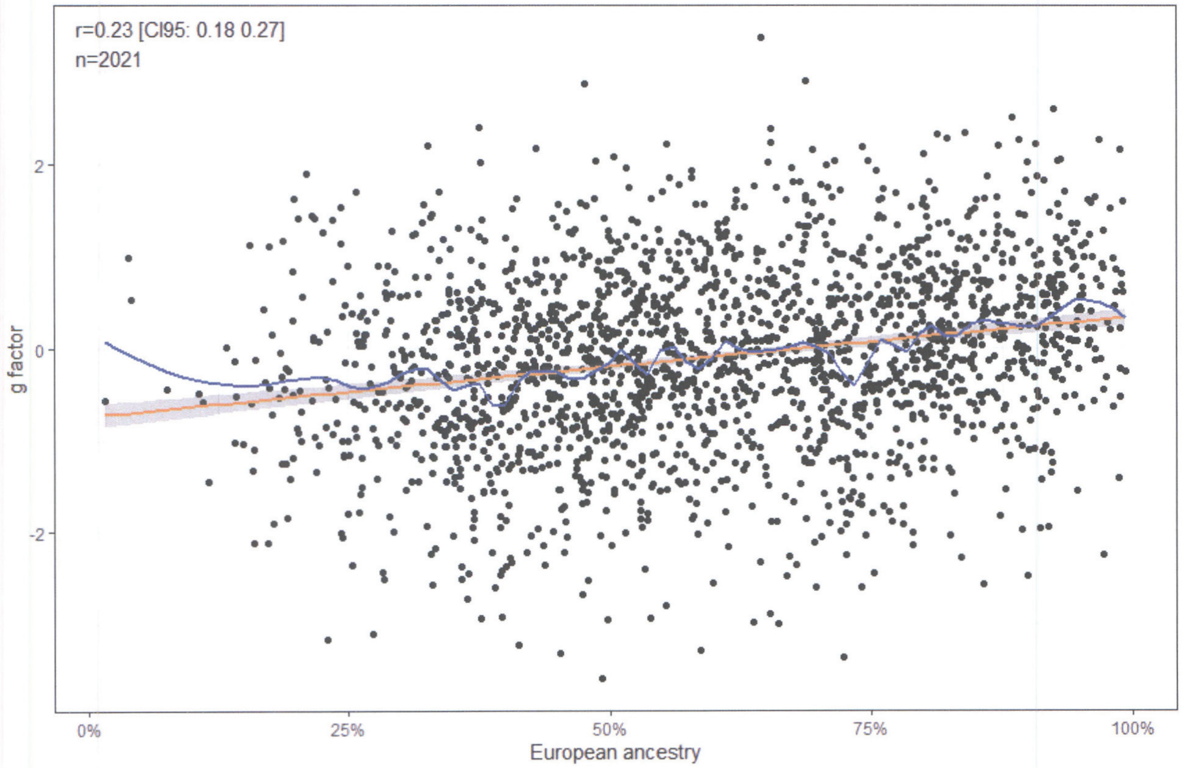


Figure 3. Regression plot of European ancestry and *g* in the Hispanic American subsample ($N = 2021$)

For the Hispanic admixture-regression analyses, we include a term for race because the Hispanic ethnic category is inclusive of all self-identified racial groups. As shown in Table 3, both Amerindian and African ancestry are strongly negatively associated with *g* in the first three models. Adding SIRE ethnic identity and the ethnic attachment variable in Model 2 had little effect on the beta for Amerindian ancestry. Doing so increases the effect of African ancestry. In Model 3, both skin color and experiences of discrimination have significant independent effects on *g*, but these variables only slightly attenuated the relation between *g* and Amerindian and African ancestry. However, as seen in Model 4, SES attenuated the effect of Amerindian and African ancestry (Model 3: $b_{\text{african ancestry}} = -0.96$ →Model 4: $b_{\text{african ancestry}} = -0.58$; Model 3: $b_{\text{amerindian ancestry}} = -1.37$ →Model 4: $b_{\text{amerindian ancestry}} = -0.86$). Nonetheless, the magnitudes of the Amerindian and African ancestry effects remained medium to large in size and statistically significant.

Table 3. Regression results for the effect of genetic ancestry on *g* among Hispanic American children ($N = 2021$). Shown are the beta coefficients (*b*) and *p*-values (*p*) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom

	Model 1		**Model 2**		**Model 3**		**Model 4**	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.31 (0.12)	**0.009**	0.34 (0.12)	**0.005**	0.09 (0.17)	0.589	-0.07 (0.17)	0.694
Amerindian_ancestry	-1.58 (0.15)	**<0.001**	-1.50 (0.15)	**<0.001**	-1.37 (0.18)	**<0.001**	-0.86 (0.18)	**<0.001**
African_ancestry	-0.76 (0.16)	**<0.001**	-1.08 (0.25)	**<0.001**	-0.96 (0.26)	**<0.001**	-0.58 (0.26)	**0.027**
East_Asian_ancestry	-1.73 (0.93)	0.062	-1.59 (0.93)	0.088	-1.57 (0.92)	0.087	-1.43 (0.90)	0.112
South_Asian_ancestry	4.23 (3.46)	0.222	4.10 (3.45)	0.235	3.83 (3.41)	0.262	3.55 (3.33)	0.287
Child_US_Born	0.09 (0.09)	0.322	0.08 (0.09)	0.408	0.08 (0.09)	0.383	0.08 (0.09)	0.364
Immigrant_Family	0.17 (0.06)	**0.003**	0.16 (0.06)	**0.005**	0.15 (0.06)	**0.009**	0.12 (0.06)	**0.026**
Mexican	-0.14 (0.06)	**0.014**	-0.16 (0.06)	**0.009**	-0.14 (0.06)	**0.017**	-0.10 (0.06)	0.090
Cuban	-0.24 (0.09)	**0.009**	-0.25 (0.09)	**0.007**	-0.23 (0.09)	**0.011**	-0.15 (0.09)	0.104
Puerto_Rican	-0.40 (0.09)	**<0.001**	-0.39 (0.09)	**<0.001**	-0.39 (0.09)	**<0.001**	-0.29 (0.09)	**0.001**
frac_Black_SIRE			0.25 (0.15)	0.107	0.28 (0.15)	0.060	0.22 (0.15)	0.144
frac_Native_American_SIRE			-0.09 (0.17)	0.599	-0.04 (0.17)	0.812	0.03 (0.16)	0.835
frac_NOC_SIRE			-0.09 (0.06)	0.138	-0.06 (0.06)	0.272	-0.03 (0.06)	0.589
Ethnic_attachment			0.06 (0.02)	**0.008**	0.06 (0.02)	**0.008**	0.04 (0.02)	**0.045**
State_racism					0.03 (0.05)	0.533	0.06 (0.05)	0.236
discrim_fact					-0.14 (0.02)	**<0.001**	-0.13 (0.02)	**<0.001**
Skin_color					-0.08 (0.04)	**0.032**	-0.07 (0.04)	**0.044**
P_Brown_Eye					0.20 (0.11)	0.072	0.26 (0.11)	**0.017**
P_Intermediate_Eye					0.20 (0.44)	0.653	0.26 (0.43)	0.548
P_Black_Hair					0.03 (0.19)	0.869	-0.03 (0.18)	0.865
P_Brown_Hair					0.12 (0.19)	0.541	0.12 (0.19)	0.535
SES							0.25 (0.03)	**<0.001**
Random Effects								
σ^2	0.40		0.39		0.39		0.39	
τ_{00}	$0.49_{\text{site\_id\_1:rel\_family\_id}}$ $0.02_{\text{site\_id\_1}}$		$0.48_{\text{site\_id\_1:rel\_family\_id}}$ $0.02_{\text{site\_id\_1}}$		$0.46_{\text{site\_id\_1:rel\_family\_id}}$ $0.03_{\text{site\_id\_1}}$		$0.42_{\text{site\_id\_1:rel\_family\_id}}$ $0.03_{\text{site\_id\_1}}$	
ICC	0.56		0.56		0.56		0.54	
N	$22_{\text{site\_id\_1}}$ $1794_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $1794_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $1794_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $1794_{\text{rel\_family\_id}}$	
Observations	2021		2021		2021		2021	
Marginal R^2 / Conditional R^2	0.087 / 0.600		0.094 / 0.603		0.119 / 0.611		0.160 / 0.611	

3.2.3. Other Americans

Figure 4 shows the regression plot for European ancestry and *g*-scores among the Other group. European ancestry is significantly ($r = .19$, $N = 748$) associated with *g* scores. Eight outliers were identified using the R boxplot function. Removing these had little effect on the correlation ($r = .16$, $N = 740$). The Loess regression line shows a slight uptick in scores at the lowest European ancestry decile. However, the 95% confidence intervals of this line overlapped with the linear regression line. The correlation matrix is provided in the Supplementary File.

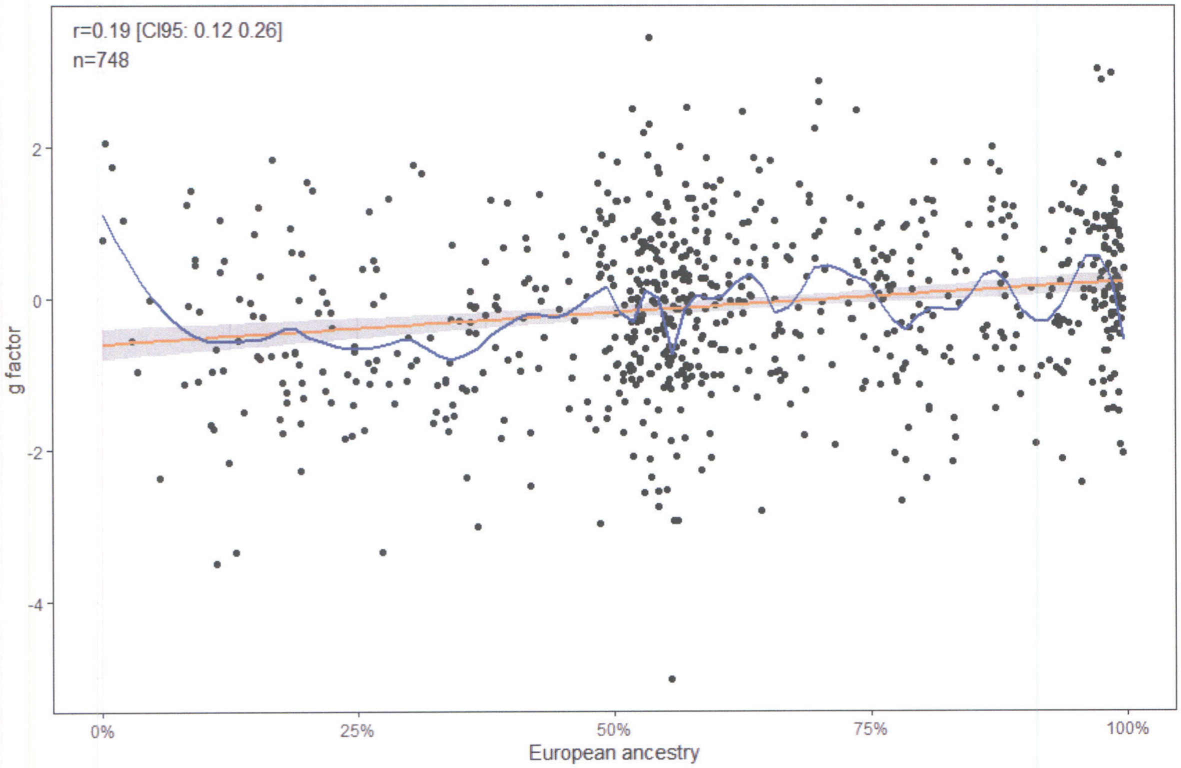


Figure 4. Regression plot of European ancestry and *g* in Other American subsample ($N = 748$)

As for Hispanics, we include a term for race because the Other American ethnic category is inclusive of all self-identified racial groups. As seen in Table 4, both coefficients for Amerindian and African ancestry show a strong negative association with *g* from Model 1 through Model 4. Adding SIRE ethnic identity and the ethnic attachment variable in Model 2 has little effect on the beta for Amerindian ancestry. Doing so increases the effect of African ancestry. In Model 3, only experiences of discrimination has a significant independent effect on *g*. The discrimination variables did not attenuate the relation between *g* and Amerindian and African ancestry. As seen in Model 4, SES moderately attenuated the effects of Amerindian and African ancestry (Model 3: $b_{\text{african ancestry}} = -1.38$ →Model 4: $b_{\text{african ancestry}} = -1.08$; Model 3: $b_{\text{amerindian ancestry}} = -1.51$ →Model 4: $b_{\text{amerindian ancestry}} = -1.09$). Nonetheless, the magnitudes of these ancestry effects remained large in magnitude and statistically significant.

Table 4. Regression results for the effect of genetic ancestry on *g* among Other Americans children ($N=748$). Shown are the beta coefficients (b) and p-values (p) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom

	Model 1		**Model 2**		**Model 3**		**Model 4**	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	-0.14 (0.32)	0.671	0.03 (0.34)	0.923	0.14 (0.39)	0.716	0.01 (0.38)	0.977
Amerindian_ancestry	-1.55 (0.45)	**0.001**	-1.40 (0.48)	**0.003**	-1.51 (0.51)	**0.003**	-1.09 (0.50)	**0.031**
African_ancestry	-1.07 (0.17)	**<0.001**	-1.27 (0.24)	**<0.001**	-1.38 (0.29)	**<0.001**	-1.08 (0.29)	**<0.001**
East_Asian_ancestry	0.95 (0.56)	0.088	0.96 (0.57)	0.093	0.59 (0.60)	0.328	0.57 (0.58)	0.330
South_Asian_ancestry	0.61 (0.79)	0.436	0.74 (0.80)	0.354	0.65 (0.81)	0.425	1.11 (0.79)	0.160
Child_US_Born	0.43 (0.30)	0.155	0.32 (0.31)	0.299	0.27 (0.31)	0.371	0.27 (0.30)	0.362
Immigrant_Family	0.14 (0.11)	0.215	0.09 (0.11)	0.433	0.07 (0.12)	0.564	-0.04 (0.11)	0.718
frac_Black_SIRE			0.16 (0.33)	0.625	0.18 (0.33)	0.584	0.22 (0.32)	0.504
frac_Native_American_SIRE			-0.22 (0.23)	0.346	-0.23 (0.24)	0.337	-0.21 (0.23)	0.362
frac_NOC_SIRE			-0.13 (0.20)	0.528	-0.15 (0.20)	0.435	-0.06 (0.19)	0.757
Ethnic_attachment			0.07 (0.04)	0.086	0.08 (0.04)	0.057	0.06 (0.04)	0.095
State_racism					0.03 (0.07)	0.683	0.07 (0.07)	0.322
discrim_fact					-0.11 (0.04)	**0.002**	-0.10 (0.04)	**0.007**
Skin_color					0.01 (0.06)	0.832	0.01 (0.06)	0.861
P_Brown_Eye					0.11 (0.17)	0.514	0.09 (0.17)	0.606
P_Intermediate_Eye					0.72 (0.75)	0.343	0.66 (0.73)	0.370
P_Black_Hair					0.22 (0.33)	0.506	0.33 (0.32)	0.315
P_Brown_Hair					-0.36 (0.31)	0.243	-0.21 (0.31)	0.493
SES							0.27 (0.05)	**<0.001**
Random Effects								
σ^2	0.62		0.61		0.62		0.65	
τ_{00}	0.45 $_{\text{site\_id\_1:rel\_family\_id}}$ 0.06 $_{\text{site\_id\_1}}$		0.46 $_{\text{site\_id\_1:rel\_family\_id}}$ 0.06 $_{\text{site\_id\_1}}$		0.44 $_{\text{site\_id\_1:rel\_family\_id}}$ 0.06 $_{\text{site\_id\_1}}$		0.35 $_{\text{site\_id\_1:rel\_family\_id}}$ 0.07 $_{\text{site\_id\_1}}$	
ICC	0.45		0.46		0.45		0.39	
N	22 $_{\text{site\_id\_1}}$ 644 $_{\text{rel\_family\_id}}$		22 $_{\text{site\_id\_1}}$ 644 $_{\text{rel\_family\_id}}$		22 $_{\text{site\_id\_1}}$ 644 $_{\text{rel\_family\_id}}$		22 $_{\text{site\_id\_1}}$ 644 $_{\text{rel\_family\_id}}$	
Observations	748		748		748		748	
Marginal R^2 / Conditional R^2	0.076 / 0.493		0.083 / 0.503		0.098 / 0.503		0.142 / 0.481	

3.2.4. White Americans

We do not report the admixture regression results for the 5911 non-Hispanic White Americans. These results are unreliable owing to the low dispersion in African and Amerindian ancestry within this SIRE group (see Table 1). Thus, we relegate these results to the supplemental material. Briefly, though, in Model 4 for this subsample, both African ancestry ($b_{\text{African ancestry}} = -.85$, $p = .110$) and Amerindian ancestry ($b_{\text{Amerindian ancestry}} = -.96$) also have large negative effects. However, this effect is only statistically significant for Amerindian ancestry ($p = .012$).

3.2.5. Full Sample

The results above indicate that factors associated with genetic ancestry are related to *g* *within* SIRE groups. These findings also suggest that these same factors explain differences *between* SIRE groups (Halder et al., 2015). Using the full sample, we examine this implication. The relation between European ancestry and *g* for the full sample is shown in Figure 5. As expected, there is a strong positive association for the SIRE groups between ancestry and *g* ($r = .36$; $N = 10370$). Because the range of ancestry is not restricted — restriction of range attenuates correlations — the correlation is high. In this plot, we again see the uptick at the lowest decile of European admixture. This is due to the relatively high scores of children of recent African immigrants.

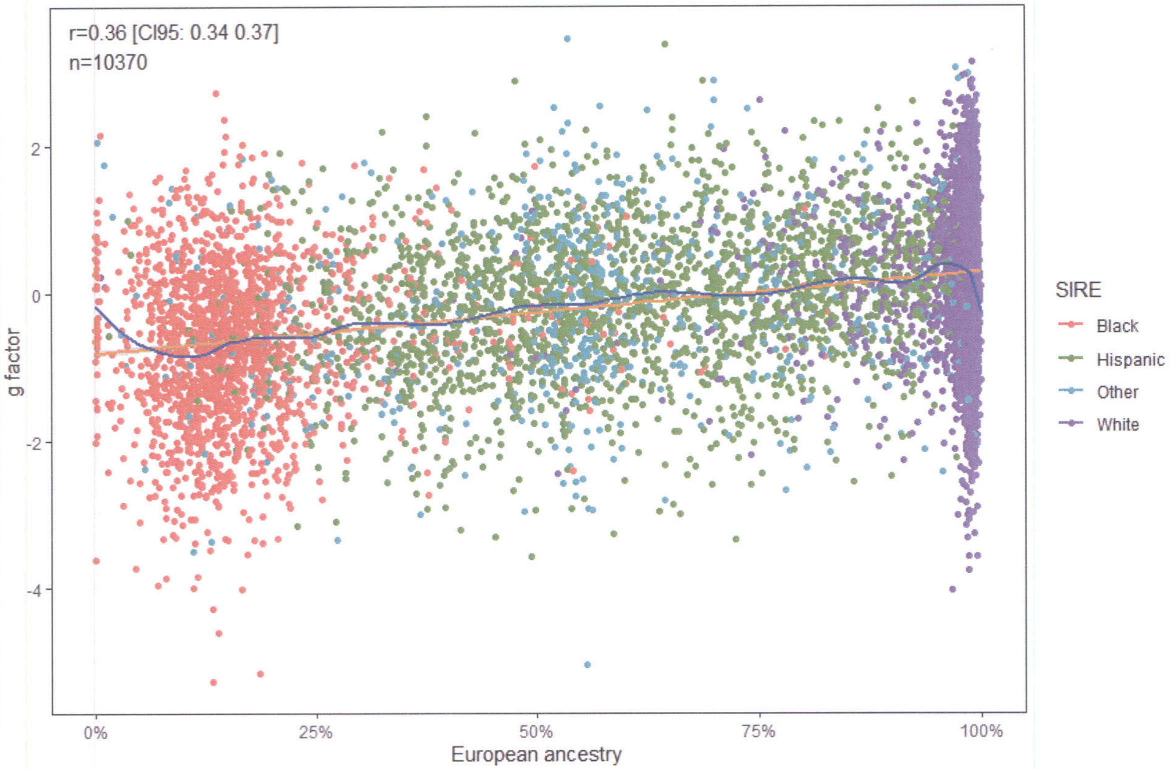


Figure 5. Regression plot of European ancestry and *g* in the full sample ($N = 10370$)

To examine if SIRE differences can be accounted for by genetic ancestry, we construct a new set of regression models using the full sample. As seen in Table 5 in the first two models, Model 1a and Model 1b, we include only genetic ancestry variables or alternatively SIRE variables along with controls for migrant status. As seen in Model 2, none of the SIRE values remain significant after adding genetic ancestry to the model. These results indicate that ancestry-associated factors account for the SIRE differences in *g*. We additionally include a Model 3, which adds the cultural, socioeconomic, and phenotypic indices. As seen in Model 3, these variables attenuated the effect of African and Amerindian ancestry (Model 2: $b_{\text{african ancestry}} = -1.31$ →Model 3: $b_{\text{african ancestry}} = -0.80$; Model 2: $b_{\text{amerindian ancestry}} = -1.57$ →Model 3: $b_{\text{amerindian ancestry}} = -0.86$), but the ancestry effects remain large. Note that the effects of East Asian and South Asian ancestry are insignificant because there is little variance in these ancestry components. This is because we excluded everyone identified as Asian and Pacific Islander.

Table 5. Regression results for the effect of ancestry on cognitive ability in the full sample (*N* = 10,370). Shown are the beta coefficients (*b*) and *p*-values (*p*) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom

	Model 1a		**Model 1b**		**Model 2**		**Model 3**	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.18 (0.07)	**0.015**	0.16 (0.07)	**0.024**	0.18 (0.07)	**0.013**	-0.02 (0.08)	0.855
Amerindian_ancestry	-1.60 (0.08)	**<0.001**			-1.57 (0.12)	**<0.001**	-0.86 (0.12)	**<0.001**
African_ancestry	-1.30 (0.03)	**<0.001**			-1.31 (0.11)	**<0.001**	-0.79 (0.11)	**<0.001**
East_Asian_ancestry	0.06 (0.33)	0.863			0.12 (0.33)	0.711	0.21 (0.32)	0.509
South_Asian_ancestry	0.35 (0.56)	0.533			0.44 (0.56)	0.437	0.84 (0.54)	0.122
Child_US_Born	0.14 (0.06)	**0.027**	0.12 (0.06)	0.059	0.13 (0.06)	**0.031**	0.17 (0.06)	**0.005**
Immigrant_Family	0.15 (0.03)	**<0.001**	0.10 (0.03)	**<0.001**	0.15 (0.03)	**<0.001**	0.09 (0.02)	**<0.001**
frac_Black_SIRE			-0.99 (0.03)	**<0.001**	0.01 (0.09)	0.914	0.04 (0.08)	0.652
frac_Native_American_SIRE			-0.38 (0.09)	**<0.001**	-0.09 (0.09)	0.333	0.00 (0.09)	0.978
frac_NOC_SIRE			-0.42 (0.05)	**<0.001**	-0.08 (0.05)	0.097	-0.01 (0.05)	0.807
Hispanic			-0.29 (0.03)	**<0.001**	0.02 (0.04)	0.623	0.06 (0.04)	0.136
State_racism							0.08 (0.04)	0.050
discrim_fact							-0.09 (0.01)	**<0.001**
Skin_color							-0.01 (0.02)	0.525
P_Brown_Eye							0.03 (0.04)	0.376
P_Intermediate_Eye							-0.07 (0.15)	0.654
P_Black_Hair							-0.05 (0.08)	0.541
P_Brown_Hair							0.03 (0.06)	0.638
SES							0.31 (0.01)	**<0.001**
Random Effects								
σ^2	0.42		0.42		0.42		0.42	
τ_{00}	0.43 $_{\text{site\_id\_1:rel\_family\_id}}$		0.46 $_{\text{site\_id\_1:rel\_family\_id}}$		0.43 $_{\text{site\_id\_1:rel\_family\_id}}$		0.35 $_{\text{site\_id\_1:rel\_family\_id}}$	
	0.03 $_{\text{site\_id\_1}}$		0.02 $_{\text{site\_id\_1}}$		0.03 $_{\text{site\_id\_1}}$		0.05 $_{\text{site\_id\_1}}$	
ICC	0.53		0.53		0.52		0.49	
N	22 $_{\text{site\_id\_1}}$		22 $_{\text{site\_id\_1}}$		22 $_{\text{site\_id\_1}}$		22 $_{\text{site\_id\_1}}$	
	8672 $_{\text{rel\_family\_id}}$		8672 $_{\text{rel\_family\_id}}$		8672 $_{\text{rel\_family\_id}}$		8672 $_{\text{rel\_family\_id}}$	
Observations	10370		10370		10370		10370	
Marginal R^2 / Conditional R^2	0.178 / 0.609		0.145 / 0.600		0.178 / 0.609		0.247 / 0.615	

3.3. Analyses with Polygenic Scores

Finally, we can check the extent to which eduPGS can explain ancestry effects. Before doing so, we verify that eduPGS are associated with *g* within each of the SIRE groups. In doing so, we include controls for the first 20 genetic principal components or, alternatively, continental ancestry (with European ancestry left as the reference). Moreover, we run the analysis both using all families and using only singleton families (i.e., families with only one child). The full results are provided in the supplementary material. The results are summarized in Table 6. As previously found, the eduPGS by *g* associations are attenuated among African Americans, but not among Hispanic and Other Americans (Fuerst et al., 2021). Nonetheless, eduPGS are significantly associated with *g* within all SIRE groups.

Table 6. Validities (*b*) of eduPGS by American SIRE groups from multilevel regression models with *g* as the dependent variable and PGS as a predictor

Controls	**Sample**	**Black**	**Hispanic**	**Other**		**White**
20 PCs	Full sample	0.17	0.27	0.32		0.26
Ancestry	Full sample	0.16	0.27	0.32		0.27
20 PCs	Singletons	0.19	0.29	0.26		0.27
Ancestry	Singletons	0.18	0.29	0.28		0.26

Note: Sample sizes for the full samples are: Black ($N = 1690$), Hispanic ($N = 2021$), Other ($N = 748$) and White ($N = 5911$). The sample sizes for the singletons subsamples are: Black ($N = 1159$), Hispanic ($N = 1516$), Other ($N = 505$), and White ($N = 3674$). All betas are statistically significant at the $p < .01$ level. Singletons = single child families.

For further validation of the PGS, we correlated the eduPGS with the *NIHTBX* summary scores which we had for all groups, including Asians. We computed mean scores for all ABCD SIRE subgroups and combinations with $N \geq 50$. There were 17 such groups. We then correlated the eduPGS with the mean subgroup test scores. This correlation came to $r = .93$. Thus, we conclude, in line with Chande et al. (2020), that "the general concordance seen between genetically inferred (predicted) phenotypic differences and the observed differences for anthropometric traits, or known prevalence differences in the case of disease traits, supports the approach taken here" (*p*. 1525-6), despite concerns raised in the literature. The regression plot is shown in Figure 6. The number of individuals in each SIRE group is represented by the size of the associated data point.

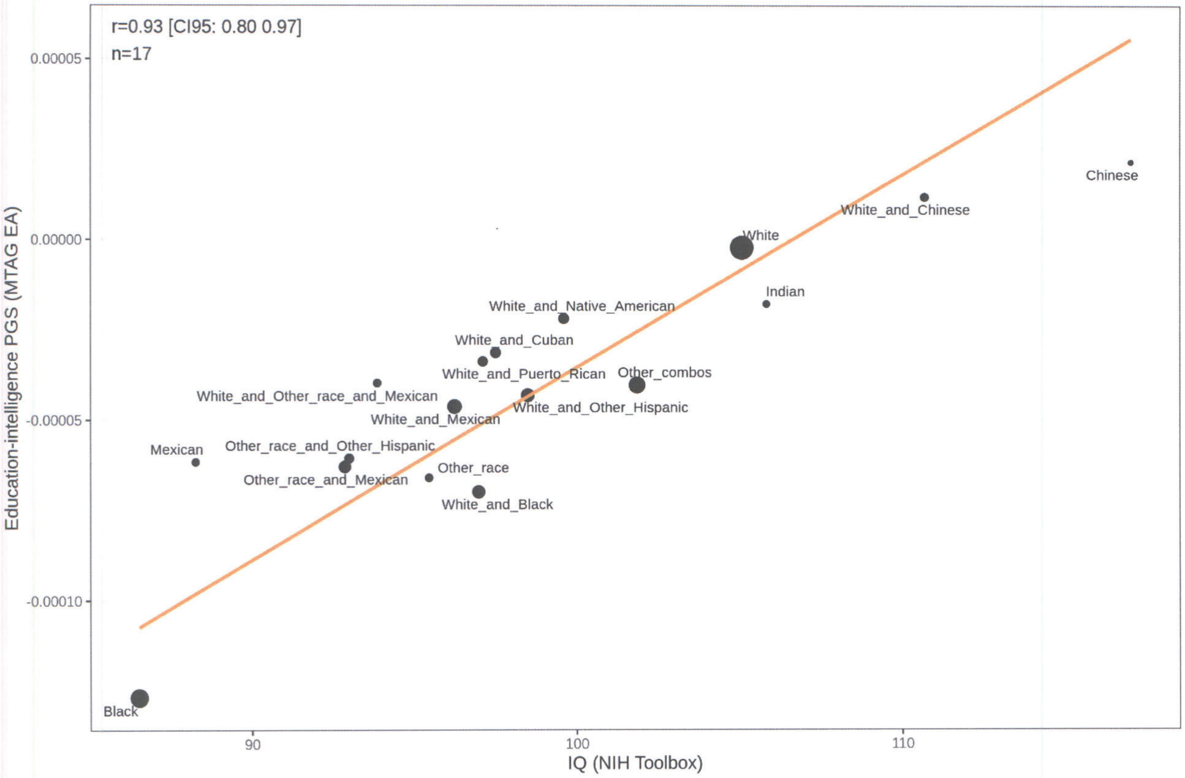
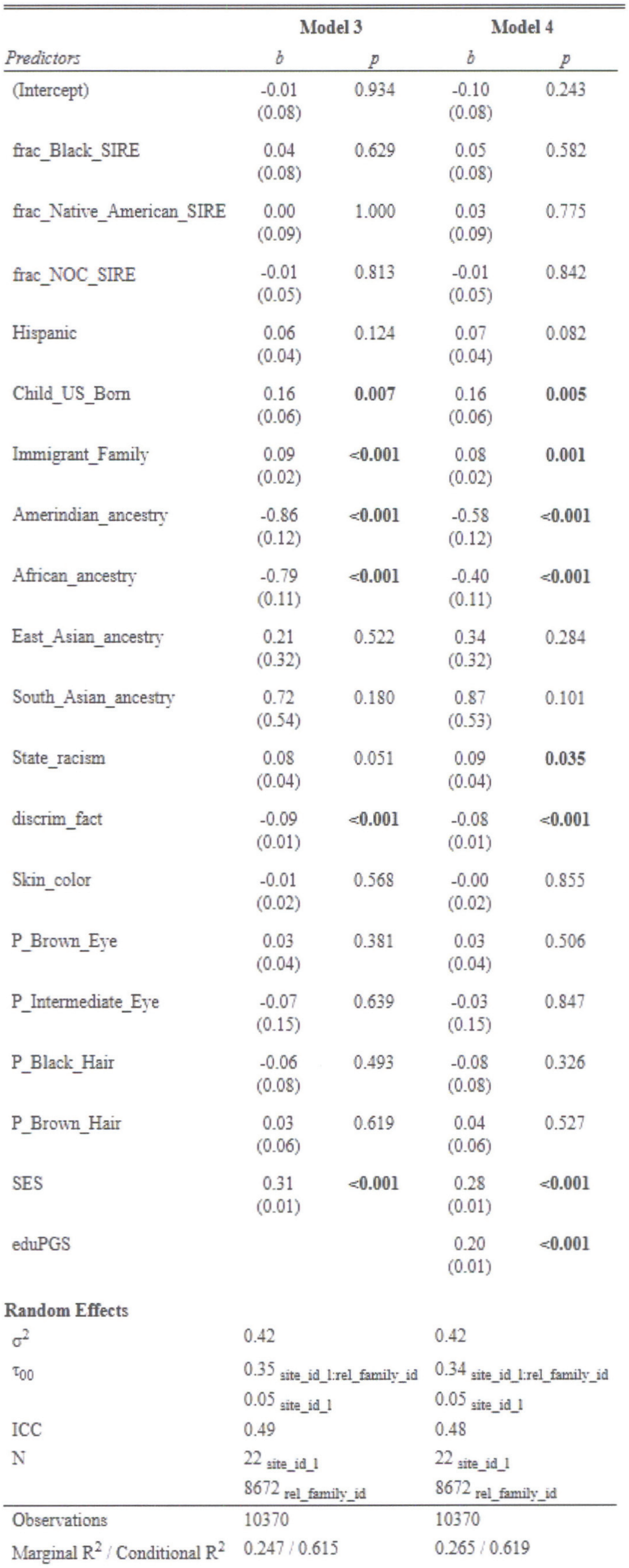


Figure 6. Regression plot of eduPGS and NIHTBX scores for the 17 largest SIRE groups in the ABCD sample, with SIRE group sample sizes represented by the size of the data points

Next, we include the PGS in the model, starting with Model 3 from Table 5. Comparing Model 3 and Model 4 (which adds eduPGS) of Table 7, we see that eduPGS explains a substantial portion of the residual effect of African and Amerindian ancestry after controls for SES.

Table 7. Regression results for the effect of eduPGS and ancestry on cognitive ability in the full sample. Shown are the beta coefficients (*b*) and *p*-values (*p*) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom



	Model 3		**Model 4**	
Predictors	*b*	*p*	*b*	*p*
(Intercept)	-0.01 (0.08)	0.934	-0.10 (0.08)	0.243
frac_Black_SIRE	0.04 (0.08)	0.629	0.05 (0.08)	0.582
frac_Native_American_SIRE	0.00 (0.09)	1.000	0.03 (0.09)	0.775
frac_NOC_SIRE	-0.01 (0.05)	0.813	-0.01 (0.05)	0.842
Hispanic	0.06 (0.04)	0.124	0.07 (0.04)	0.082
Child_US_Born	0.16 (0.06)	**0.007**	0.16 (0.06)	**0.005**
Immigrant_Family	0.09 (0.02)	**<0.001**	0.08 (0.02)	**0.001**
Amerindian_ancestry	-0.86 (0.12)	**<0.001**	-0.58 (0.12)	**<0.001**
African_ancestry	-0.79 (0.11)	**<0.001**	-0.40 (0.11)	**<0.001**
East_Asian_ancestry	0.21 (0.32)	0.522	0.34 (0.32)	0.284
South_Asian_ancestry	0.72 (0.54)	0.180	0.87 (0.53)	0.101
State_racism	0.08 (0.04)	0.051	0.09 (0.04)	**0.035**
discrim_fact	-0.09 (0.01)	**<0.001**	-0.08 (0.01)	**<0.001**
Skin_color	-0.01 (0.02)	0.568	-0.00 (0.02)	0.855
P_Brown_Eye	0.03 (0.04)	0.381	0.03 (0.04)	0.506
P_Intermediate_Eye	-0.07 (0.15)	0.639	-0.03 (0.15)	0.847
P_Black_Hair	-0.06 (0.08)	0.493	-0.08 (0.08)	0.326
P_Brown_Hair	0.03 (0.06)	0.619	0.04 (0.06)	0.527
SES	0.31 (0.01)	**<0.001**	0.28 (0.01)	**<0.001**
eduPGS			0.20 (0.01)	**<0.001**
Random Effects				
σ^2	0.42		0.42	
τ_{00}	0.35 $_{\text{site\_id\_l:rel\_family\_id}}$ 0.05 $_{\text{site\_id\_l}}$		0.34 $_{\text{site\_id\_l:rel\_family\_id}}$ 0.05 $_{\text{site\_id\_l}}$	
ICC	0.49		0.48	
N	22 $_{\text{site\_id\_l}}$ 8672 $_{\text{rel\_family\_id}}$		22 $_{\text{site\_id\_l}}$ 8672 $_{\text{rel\_family\_id}}$	
Observations	10370		10370	
Marginal R^2 / Conditional R^2	0.247 / 0.615		0.265 / 0.619	

We also examine the individual SIRE subsample results for eduPGS. The model adds eduPGS to the respective Model 4s for each SIRE group (i.e., the model with potential environmental factors included). The full results are provided in the Supplementary File. These results are summarized in Table 8. Specifically, Table 8 shows the effects for Amerindian and African ancestry on *g* with possible environmental controls. These come from the fourth models of Tables 2, 3, 4, and S8 and the third model from Table 5. It next shows the effects when eduPGS is added. As seen, eduPGS accounts for a portion of the ancestry by *g* association in all SIRE subsamples.

Table 8. Effects (*b*) of Amerindian and African ancestry on *g* in multi-level models with environmental controls (Model 4/3), and multi-level models with environmental controls and eduPGS (Model 5)

	Black		Hispanic		Other		White		Full	
	Model 4	Model 5	Model 4	Model 5	Model 4	Model 5	Model 4	Model 5	Model 3	Model 5
Amerindian	-3.46	-3.15	-0.86	-0.55	-1.09	-0.55	-0.96	-0.65	-0.86	-0.58
African	-0.89	-0.69	-0.58	-0.16	-1.08	-0.48	-0.85	-0.32	-0.79	-0.40

It is conceptually possible that our eduPGS are just capturing global ancestry effects. Our ancestry components are based on more SNPs. Moreover, they are not weighted by trait-associations which will attenuate the association with ancestry. As such this is unlikely. However, to test this possibility we created pseudoPGS. To do so, we used PLINK v1.90b6.8 to select random sets of 8,898 variants to match the eduPGS. Then we randomly assigned the eduPGS beta weights (from Lee et al., 2018) to the respective sets of SNPs.

Following this procedure, we create 10 pseudo eduPGS scores. This procedure produced PGS with the same set of SNPs as the SNPs used to calculate genetic ancestry, but randomized trait-association information. The full results are provided in Supplementary File 3, Table S17. Unlike the real PGS, these pseudoPGS had no validity independent of genetic ancestry. This is because of the random assignment of eduPGS betas to the SNP frequencies resulting in poor indices of ancestry. Generally, we conclude that PGS will not necessarily capture effects of global ancestry. This finding suggests that our eduPGS are in fact capturing causal genetic effects on *g* both within and between ancestries.

4. Conclusion

4.1. Discussion

Genetic ancestry measures provide very powerful scientific value in studying SIRE differences in *g*. Using ancestry allows one to examine how the trait varies by genetic ancestry within self-identified racial and ethnic groups. Doing so offers a potential solution to the problem of decomposing genetic and environmental variance (Halder et al., 2015). Admixture regression has been widely applied to medical and behavioral traits. This includes type 2 diabetes (Cheng et al., 2013), asthma (Salari et al., 2005), blood pressure (Klimentidis et al., 2012), and sleep depth (Halder et al., 2015). Admixture regression has a natural application to studying *g*.

Here we apply this technique to examine SIRE differences in *g*. We find that African and Amerindian ancestry are strongly negatively associated with general cognitive ability among African, Hispanic, and other American subsamples. This replicates previous research which showed that genetic ancestry predicts cognitive ability, independent of socioeconomic status and phenotypic discrimination variables which are the usual suspects (Kirkegaard et al., 2019; Lasker et al., 2019; Warne, 2020b). The importance of such analyses *within* SIRE groups is that they shed light on the cause of *g* differences *between* SIRE groups with respect to similarities in developmental processes (Rowe, Vazsonyi & Flannery, 1994).

The ancestry effects are consistent in direction across subsamples and hold after controlling for a wide array of economic and social factors, including migrant status, SIRE, ethnic attachment, measures of discrimination, phenotypic indices of race, and general SES. These results suggest that African, Hispanic, and other groups have inherited alleles from their African and Amerindian ancestors which make them liable to lower levels of *g*. In fact, as seen in Table 5 (Model 2), 100%, 76%, 81%, and 100% of the respective Black, Native American, Other, and Hispanic SIRE effects were explained by genetic ancestry. This association between genetic ancestry and *g* suggests a partial genetic basis for observed SIRE differences.

This inference is supported by additional findings based on the eduPGS analyses. These polygenic scores were found to be predictive of *g* within SIRE groups controlling for the first 20 principal components and for ancestry. Moreover, they explain a substantial portion of the ancestry effects both in the full sample and all subsamples. Also, they were almost perfectly correlated with SIRE group means in cognitive ability ($r = .93$). The most parsimonious explanation for this, given the apparent absence of obvious forms of confounding (Fuerst et al., 2021), would seem to be that eduPGS are capturing causal effects of genes on *g* both within and between ancestry groups and thus also SIRE groups. Firm conclusions, though, will require a better understanding of the relation between polygenic scores and ancestry (Lawson et al., 2020; Fuerst et al., 2021).

It is worth emphasizing that our *g* scores were from a confirmatory factor model in which strict factorial invariance (SFI) held between SIRE groups. SFI entails that the differences between SIRE groups have the same psychometric meaning as the differences between individuals within these groups (i.e., the scores are psychometrically unbiased). Moreover, SFI implies that the causes of group differences are a subset of the causes of the individual differences within groups (Lubke et al., 2003; Dalliard, 2014). In this sample of children, individual differences in general cognitive ability are largely due to genes (Freis et al., 2020).

It should be noted that the polygenic scores represent genetic variation that is caused by common alleles, not genetic variation that is caused by rare alleles under mutation-selection balance. The causal alleles that are tapped by polygenic scores are ancient. Most were already polymorphic 60,000 years ago when people left Africa and spread all over Eurasia. Today's racial allele frequency differences are the cumulative effects of selection and genetic drift acting over more than 2,000 generations, while rare variants under mutation-selection balance are much younger, no more than one or two millennia or even less. Therefore it is predictable that genetic race differences that evolved over a long time are differences in polygenic scores but not necessarily differences in mutational load. The latter are the result of strength of selection during the last centuries.

Overall, the results suggest that genetic variants related to general cognitive ability vary between source genetic populations and have a causal effect on intelligence. Because individuals within SIRE groups differ in their proportion of African, European and Amerindian ancestors, general cognitive ability varies by genetic ancestry within SIRE groups.

4.2. Limitations

This study advances over previous studies in that we used a diverse national sample, a good measure of *g*, multiple indices of racial discrimination including multiple race-associated phenotypes, and a composite index of SES based on seven different indices. Moreover, our multilevel model controlled for the effects of geography. Unfortunately, our index of skin color was imperfect. However, it seems unlikely that skin color discrimination is a significant *immediate* cause of *g* differences among 9-10 year old children. Such color discrimination explanations usually propose labor market based discrimination (Hersch, 2011), which would be captured by our index of SES. Regardless, admixture-regression results can only provide indirect evidence for a genetic hypothesis because there could be unmeasured environmental factors that are related to both ancestry and cognitive ability.

While the results also show that educational and intelligence-related polygenic scores can account for some of the effects of ancestry on *g*, these results are only tentative. It is not certain that these PGS are capturing genetic effects, at least between ancestries (Fuerst et al., 2021). Thus these results do not provide definitive evidence for a genetic hypothesis. However, following the methodology of genetic epidemiology, admixture regression analyses are just a first step in elucidating the genetic and environmental causes of group differences.

Author contributions

All analyses were conducted by JGRF. MH and GC helped prepare the manuscript.

Acknowledgments

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data, but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the research results and interpretations of the authors alone and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Version 3.01. The raw data are available at https://nda.nih.gov/edit_collection.html?id=2573. Instructions on how to create an NDA study are available at https://nda.nih.gov/training/modules/study.html). Additional support for this work was made possible from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680), and Children and Screens: Institute of Digital Media and Child Development Inc.

Additionally, we thank Jordan Lasker for guidance with the MGCFA analysis code.

References

Belsky, D.W., Domingue, B.W., Wedow, R., Arseneault, L., Boardman, J.D., Caspi, A., ... & Harris, K.M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences* 115: E7275-E7284.

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics* 96: 37-53.

Chande, A.T., Rishishwar, L., Ban, D., Nagar, S.D., Conley, A.B., Rowell, J., ... & Jordan, I.K. (2020). The phenotypic consequences of genetic divergence between admixed Latin American populations: Antioquia and Chocó, Colombia. *Genome Biology and Evolution* 12: 1516-1527.

Cheng, C.Y., Reich, D., Haiman, C.A., Tandon, A., Patterson, N., Elizabeth, S., ... & Kao, W.L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: A genetic admixture analysis in three U.S. population cohorts. *PLoS One* 7(3): e32840.

Choi, S.W., Mak, T.S.H. & O'Reilly, P.F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols* 15: 2759-2772.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Abingdon, UK: Routledge.

Conley, D. & Fletcher, J. (2017). *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton Univ. Press: Princeton, NJ.

Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J. (2014). Multivariate analysis of mixed data: The R Package PCAmixdata. *arXiv preprint arXiv:1411.4911*. https://arxiv.org/pdf/1411.4911.pdf

Dalliard, M. (2014). The elusive x-factor: A critique of J.M. Kaplan's model of race and IQ. *Open Differential Psychology*: 1-44.

Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., ... & DuVall, S.L. (2019). Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *American Journal of Human Genetics* 105: 763-772.

Freis, S., Morrison, C., Lessem, J.M., Hewitt, J.K. & Friedman, N.P. (2020). Genetic and environmental influences on executive functions and intelligence in the ABCD study. PsyArXiv. https://psyarxiv.com/xpba4/

Fuerst, J.G.R., Kirkegaard, E.O., Piffer, D. (2021). More research needed: There is a robust causal vs. confounding problem for intelligence-associated polygenic scores in context to admixed American populations. *Mankind Quarterly*, this issue.

Halder, I., Matthews, K.A., Buysse, D.J., Strollo, P.J., Causer, V., Reis, S.E. & Hall, M.H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep* 38: 1185-1193.

Heeringa, S.G. & Berglund, P.A. (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. *BioRxiv*.

Hersch, J. (2011). The persistence of skin color discrimination for immigrants. *Social Science Research* 40: 1337-1349.

Hu, M., Lasker, J., Kirkegaard, E.O. & Fuerst, J.G. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych* 1(1): 240-261.

Kim, M.S., Patel, K.P., Teng, A.K., Berens, A.J. & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome Biology* 19(1): 1-14.

Kirkegaard, E.O., Williams, R.L., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych* 1(1): 1-25.

Klimentidis, Y.C., Dulin-Keita, A., Casazza, K., Willig, A.L., Allison, D.B. & Fernandez, J.R. (2012). Genetic admixture, social-behavioural factors and body composition are associated with blood pressure differently by racial-ethnic group among children. *Journal of Human Hypertension* 26: 98-107.

Krapohl, E. & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry* 21: 437-443.

Lasker, J., Pesta, B.J., Fuerst, J.G. & Kirkegaard, E.O. (2019). Global ancestry and cognitive ability. *Psych* 1(1): 431-459.

Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., ... & Timpson, N.J. (2020). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* 139: 23-41.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & 23andMe Research Team (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics* 50: 1112-1121.

Liebler, C.A. & Halpern-Manners, A. (2008). A practical approach to using multiple-race response data: A bridging method for public use microdata. *Demography* 45: 143-155.

Lubke, G.H., Dolan, C.V., Kelderman, H. & Mellenbergh, G.J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence* 31: 543-566.

Magnuson, K.A. & Duncan, G.J. (2006). The role of family socioeconomic resources in the black–white test score gap among young children. *Developmental Review* 26: 365-399.

Murray, C. (2021). *Facing Reality: Two Truths about Race in America.* Encounter Books.

Pesta, B.J., Kirkegaard, E.O., te Nijenhuis, J., Lasker, J. & Fuerst, J.G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence* 78: 101408.

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from GWAS hits: A replication of previous findings using recent data *Psych* 1(1): 55-75.

Rindermann, H., Becker, D. & Coyle, T.R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence* 78: 101406.

Roth, P.L., Van Iddekinge, C.H., DeOrtentiis, P.S., Hackney, K.J., Zhang, L. & Buster, M.A. (2017). Hispanic and Asian performance on selection procedures: A narrative and meta-analytic review of 12 common predictors. *Journal of Applied Psychology* 102: 1178-1202.

Rowe, D.C., Vazsonyi, A.T. & Flannery, D.J. (1994). No more than skin deep: Ethnic and racial similarity in developmental processes. *Psychological Review* 101: 396-413.

Rowe, D.C., Vesterdal, W.J. & Rodgers, J.L. (1998). Herrnstein's syllogism: Genetic and shared environmental influences on IQ, education, and income. *Intelligence* 26: 405-423.

Salari, K., Choudhry, S., Tang, H., Naqvi, M., Lind, D., Avila, P.C., ... & Ziv, E. (2005). Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genetic Epidemiology* 29: 76-86.

Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B. & Coleman, J.R. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics* 50: 912-919.

Scheiber, C. (2016a). Do the Kaufman tests of cognitive ability and academic achievement display construct bias across a representative sample of Black, Hispanic, and Caucasian school-age children in grades 1 through 12? *Psychological Assessment* 28: 942-952.

Scheiber, C. (2016b). Is the Cattell–Horn–Carroll-based factor structure of the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) construct invariant for a representative sample of African-American, Hispanic, and Caucasian male and female students ages 6 to 16 years? *Journal of Pediatric Neuropsychology* 2(3): 79-88.

Thompson, W.K., Barch, D.M., Bjork, J.M., Gonzalez, R., Nagel, B.J., Nixon, S.J. & Luciana, M. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Developmental Cognitive Neuroscience* 36: 100606.

Warne, R.T. (2020a). *In the Know: Debunking 35 Myths about Human Intelligence*. Cambridge Univ. Press.

Warne, R.T. (2020b). Continental genetic ancestry source correlates with global cognitive ability score. *Mankind Quarterly* 60: 400-422.

Weiss, L.G. & Saklofske, D.H. (2020). Mediators of IQ test score differences across racial and ethnic groups: The case for environmental and social justice. *Personality and Individual Differences* 161: 109962.

CHAPTER 10

A GENETICALLY-INFORMED TEST OF THE COGNITIVE-COLORISM HYPOTHESIS

VLADIMIR SHIBAEV AND JOHN G.R. FUERST

Abstract

In the Americas, less European-looking people have on average worse academic outcomes than more European-looking people. According to the colorism model, these associations between race-related phenotype and academic-related outcomes are due to contemporary phenotypic-based discrimination and not due to family-background or intergenerational factors. Previous studies have attempted to use sibling designs to disentangle the latter two causes from the effects of discrimination. We argue that admixture-regression analysis is an additional helpful tool for disentangling the various causes. Using a large, genetically-informed dataset, we created a genetically-based predictor of European appearance. We tested the hypothesis that European appearance will be associated with academic outcomes independent of genetic ancestry. We also tested the hypothesis that *g* mediated the relations between ancestry/European appearance and grades. We did not find evidence of this in the case of *g* (and most cognitive tests), but we did find tentative evidence in the case of parent-reported grades. When genetic ancestry was included in the models, European appearance was not significantly related to *g*. We also found that while *g* was a substantial and statistically significant mediator of the association between European ancestry and grades, this was not the case in the context of European appearance and grades. These results are in line with the position that cognitive inequalities in the US are intergenerationally transmitted, and are not the result of contemporaneous color-based discrimination. The admixture-regression method employed here could be applied to different outcomes to test for evidence of phenotypic-based discrimination or, at least, family-background independent effects.

Keywords: IQ, colorism, admixture study, genetic ancestry, *g*, school grades, discrimination

1. Introduction

In the Americas, research shows that less European-looking and/or darker-looking people are on average less successful than more European-looking and/or lighter-looking people. For example, stereotypically race-related phenotypes, such as skin, hair, and eye color, are generally associated with better socioeconomic outcomes (Hochschild, & Weaver, 2007; Hunter, 2013). Many sociologists attribute this association to phenotypic-based discrimination, also known as "colorism" (e.g., Dixon & Telles, 2017). According to theorists of colorism, phenotypic-based discrimination is common in the Americas, favoring individuals with stereotypical European phenotypes. This discrimination is hypothesized to result in better health, higher occupational and educational attainment, and other more positive outcomes for individuals with a paler or 'whiter' appearance;

even among those categorized as belonging to the same race/ethnicity, the paler ones will do better than the darker ones (Hochschild & Weaver, 2007; Marira & Mitra, 2013).

Proponents of this colorism model frequently maintain that 'race' is a social construct according to which people are arbitrarily categorized based on geographic origins and physical appearance and that modern racial categories have originated primarily based on the work of eighteenth- and nineteenth-century European naturalists and anthropologists (Dixon & Telles, 2017). 'Color,' in contrast, is not a social construct but refers to gradations of physical appearance associated with different levels of skin tone (Dixon & Telles, 2017). Theorists of the colorism model place primacy on "the causal role of skin tone in engendering the colorism phenomenon" (Marira & Mitra, 2013, *p*. 103). Many theorists of colorism argue that other visually conspicuous, stereotypically race-associated traits such as hair color, eye color, hair texture, and facial features may also elicit phenotypic discrimination (Crutchfield et al., 2022; Ryabov, 2013).

In addition to differences in income, health outcomes, and occupational status, colorism has been invoked to explain academic and cognitive differences, including differences in attained years of education, grade point average, and academic/cognitive test scores (Hailu, 2018; Hill, 2002; Kim & Calzada, 2019; Liu et al., 2022; Thompson & McDonald, 2016). In their review, Crutchfield et al. (2022, *p*. 10) conclude that "lighter skin tones and more Eurocentric features were linked to better academic outcomes, including higher GPAs, additional years of schooling, [and] improved academic performance".

Concerning specific mechanisms by which individual variability in color could be linked to variation in academic and cognitive outcomes, Thompson and McDonald (2016, *p*. 6) argue for discrimination, including "direct mechanisms—educational encouragement, evaluation, and provision of learning opportunities—as well as indirect mechanisms—the use of disciplinary actions". Crutchfield et al. (2022, *p*. 10) similarly emphasize teacher-student relations as causes, further noting that "darker-skinned students face the greatest barriers to optimal educational outcomes due to differential treatment". Similarly, Hannon (2014) suggests that adults and educators may have a light-skin-equals-intelligence bias which, in turn, influences both their expectations and their treatment of children of different complexion.

According to theorists of colorism, appearance-based outcome differences directly result from appearance-based discrimination. Ancestry or racial identification matters in so far as "racial classifications are determined more closely by how one phenotypically appears to belong to one race rather than strictly by one's ancestors" (Hernández, 2015, *p*. 684), and the same author argues that this is especially true in parts of Latin America. As Hall (2020, *p*. 79) notes, "in consideration of racism as pertains to colorism, ultimately such biological attributes as ancestry and bloodline may be all but completely irrelevant in the course of discrimination via various acts of colorism". Making a related point, Harris (2008, *p*. 61) states that "[t]raditional racism places a higher value on ancestry than colorism… while colorism assigns people to places along a spectrum from dark to light, indigenous or African to European". Thus, from the perspective of the proponents of colorism, ancestry per se is irrelevant.

If color phenotypes are found to merely proxy the effects of genetic ancestry, the results would be more consistent with what Abascal and Garcia (2022) describe as the inherited (dis)advantage model or what Hu et al. (2019) name the distributional model, and which we consider essentially the same model. According to this model, for various reasons, populations differ in traits, and these trait differences are transmitted vertically or intergenerationally. Because both racial appearance/color and family heritage, including ancestry, correlate in ancestrally heterogeneous populations, there is potential confounding between ancestry-related family influences and discrimination conditioned on color phenotypes.

Due to a concern for confounding, some research has attempted to control for intergenerational factors by employing sibling designs and measuring academic or educational outcomes (Bucca, 2018; Francis-Tan, 2016; Francis & Tannuri-Pianto, 2012; Hu et al., 2019; Kizer, 2017; Marteleto & Dondero, 2016; Mill & Stein, 2016; Rangel, 2015; Ryabov, 2016; Telles, 2004). The reasoning is that in recently-admixed populations, siblings may differ noticeably in race-associated phenotypes such as skin color (see: e.g., Leite et al., 2011), but siblings will exhibit little differences in genetic ancestry and no difference in family environment. As a result, with a sibling design, it should be possible to disentangle appearance-based effects from family-heritage-based ones. The academic outcome differences examined in these studies include the following variables: educational attainment (Buca, 2018; Francis-Tan, 2016; Kizer, 2017; Marteleto & Dondero, 2016; Mill & Stein, 2016; Rangel, 2015; Ryabov, 2016), grade-point average (Francis & Tannuri-Pianto, 2012), age-appropriate grade (Telles, 2004), aptitude test scores (Francis & Tannuri-Pianto, 2012; Hu et al., 2019), and literacy (Mill & Stein, 2016).

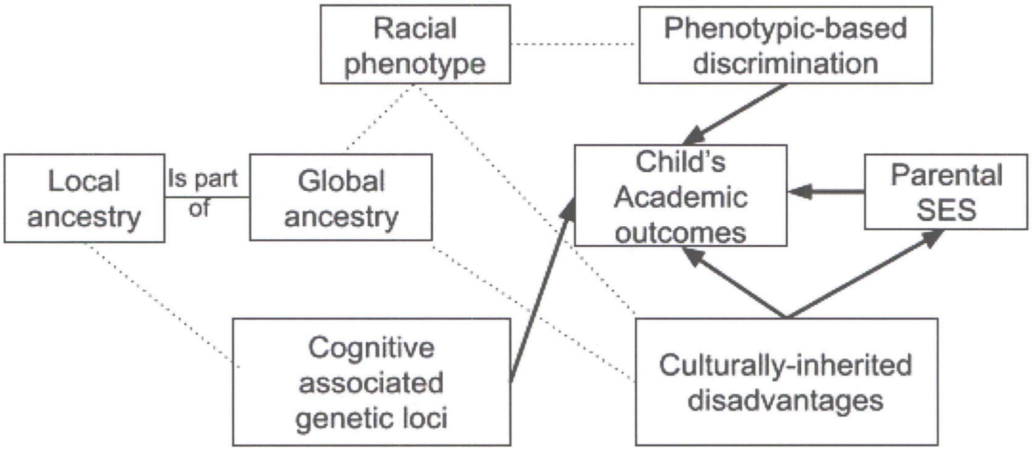
Most researchers either found modest associations between color phenotype and academic outcomes among siblings and interpreted their results as mainly supporting an intergenerational model (Francis-Tan, 2016; Mill & Stein, 2016; Rangel, 2015) or reported slight within-sibship differences but interpreted these as support for some color-based discrimination (Telles, 2004). However, a minority of researchers reported substantial and statistically significant within-sibship effects (Marteleto & Dondero, 2016; Ryabov, 2016). Although no formal meta-analysis has been conducted, carefully studying all the outcomes shows a relatively small overall within-sibships effect. The results from the eight sibling studies involving academic outcomes are reviewed in Tab 15 of the supplementary file. As seen in 16 out of 20 effects, darker siblings have on average worse academic outcomes, and the mean effect between siblings is about 10% of the effect, unconditioned on family background, among families or in the population. These analyses, though, have been limited by the modest number of variables available in the case of the census-based studies (e.g., Francis-Tan, 2016; Mill & Stein, 2016; Rangel, 2015; Telles, 2004) or non-optimal statistical power in the case of the longitudinal studies (e.g., Buca, 2018; Hu et al., 2019; Kizer, 2017).

An alternative approach to testing colorism vs. inheritance hypotheses involves using admixture-regression designs (e.g., Connor & Fuerst, 2023; Fuerst, 2021; Fuerst et al., 2021). In these designs, recently-admixed populations are treated as natural experiments, and this admixture is used to disentangle various cultural, environmental, and genetic factors. In the present context, the design is employed to try to disentangle inherited (dis)advantages associated with racial appearance from non-inherited (dis)advantages due to, for example, contemporaneous discrimination. For this purpose, global genetic ancestry and either a phenotype or genetic markers of a phenotype are included in a regression analysis along with other variables. The objective is to determine whether racial appearance affects outcomes independent of global genetic ancestry.

A theoretical path model is presented in Figure 1. The colorism model posits that discrimination based on phenotype results in disparities in cognitive abilities and other academic outcomes through the direct effects of discrimination on learning. According to this model, in populations with a mix of ancestries, academic performance is likely indirectly linked to genetic ancestry due to the correlation between racial phenotypes and genetics. Conversely, the inherited disadvantage model argues that genetic and cultural factors, which are tied to global ancestry, drive academic outcome disparities. Because genetic ancestry is correlated with racial phenotypes, academic related traits will tend to be indirectly correlated with racial phenotype in admixed populations. In the admixture-regression design, global genetic ancestry is controlled for to account for inherited disadvantages.

The colorism model clearly predicts that racial appearance affects outcomes independent of genetic ancestry. If the model is correct, individuals with lighter skin, hair, and eye color should have better outcomes controlling for overall genetic ancestry. An opposite finding would be consistent with an intergenerational model, according to which traits conducive to better outcomes are being transmitted down lines of descent as indexed by genetic ancestry. This inheritance of traits could be mediated by genes, epigenetic factors, or culturally-inherited factors. Applying the admixture-regression design to data from the US and Brazil suggests that educational attainment and cognitive ability are primarily related to genetic ancestry, and are mostly unrelated to racial appearance (Fuerst et al., 2021; Kirkegaard et al., 2017; Lasker et al., 2019).

Figure 1. Theoretical Model of the Association between Putative Causes (Discrimination vs. Inherited Disadvantage), Ancestry, Racial-Phenotype, Academic Outcomes, and Parental - Socioeconomic Status


Racial phenotype
Phenotypic-based discrimination
Local ancestry
Is part of
Global ancestry
Child's Academic outcomes
Parental SES
Cognitive associated genetic loci
Culturally-inherited disadvantages

The admixture-regression design comes with a couple of assumptions that can be easily tested. The first assumption is that there is little cross-assortative mating for race-associated phenotype and the relevant traits (e.g., cognitive ability). In the presence of such assortment, race-associated phenotype can become genetically correlated with the outcomes independent of ancestry (Jensen, 1998). If so, race-associated phenotype and outcomes may be associated with one another, independent of ancestry, for genetic and not discriminatory reasons. The second assumption is that there is no substantial reverse causation from outcomes to race-associated phenotype; for example, low-prestige occupations often are associated with increased outdoor work, sun exposure, and, consequently, darker color. These two assumptions are common to colorism research, but, in this design, they only become a concern if an association between race-associated phenotype and outcomes is found independent of genetic ancestry.

The third assumption is that genetic ancestry and racial appearance are not collinear in samples. Dissociation between race-associated traits, especially non-highly polygenic ones, and genetic ancestry is expected in admixed populations due to genetic crossover and segregation (Kim et al., 2021). The extent of dissociation is an empirical question. The final assumption is that inherited (dis)advantages correspond with differences in ancestry. This is expected based on a simple model of vertical transmission of inherited traits given an initial inequality between groups and assuming that the estimated ancestry corresponds with the percentage of ancestors of relevant parental groups. In admixed American groups this assumption holds, because genetic ancestry percentages can be understood in terms of the number of ancestors from different ancestry groups (e.g., Mooney et al., 2022).

Most studies on the effects of racial appearance use single phenotypic measures of appearance. However, interviewer-rated color scales, in particular, have often been found to have modest reliability (Campbell et al., 2020; Hannon & DeFina, 2016; Hannon & DeFina, 2020). Moreover, skin color ratings have been found to be influenced by interviewer-related characteristics (Campbell et al., 2020; Cernat et al., 2019) and, additionally, the interviewer's perceptions about the participants' socioeconomic status (Roth et al., 2022). So, there are some concerns about the reliability and validity of skin color ratings. An alternative approach is to use genetic predictors of phenotype; they have an advantage in that they do not suffer from the problems of reverse causality and interviewer-related biases. Therefore, in this study, we created several new genetically-based predictors of European appearance. Moreover, we combine these predictors of skin, hair, and eye color through factor analysis to increase reliability. Afterward, we conducted admixture-regression analyses to test if European appearance was associated with general intelligence (*g*) and school grades independently of global genetic ancestry, as predicted by theorists of colorism. Cognitive ability was of particular interest because it has been found to partially mediate the relation between color and other outcomes (Campos-Vazquez & Medina-Cortina, 2019; Fuerst et al., 2019; Kreisman & Rangel, 2015). Therefore, cognitive ability differences may partially explain the relationship between color and socioeconomic or academic outcome differences. For this reason, Huddleston and Montgomery (2010, *p*. 69) note that "more research is needed in intragroup differences among Blacks and intelligence… Results from this research have huge implications for the skin tone hierarchy in the African American community". In line with this recommendation, we examine the extent to which cognitive ability plays a mediating role for genetic ancestry and grades.

As we already noted, studies which have used sibling designs overall show a modest association between European appearance and academic outcomes. Based on these results, we hypothesized that European appearance will show an association with both *g* and grades in regression models which also include genetic ancestry. Moreover, meta-analyses indicate that general mental ability and school grades are highly correlated (e.g., Roth et al., 2015), so we hypothesized that the relation between both European appearance and genetic ancestry and between school grades will be strongly mediated by *g*.

2. Data and methods

2.1. Sample

The Adolescent Brain Cognitive Development Study (ABCD) is a collaborative longitudinal project involving 21 collection sites across the US. It was created to research the psychological and neurobiological bases of human development. At baseline, around 11,000 9-10-year-old children were sampled, mostly from public and private elementary schools. A probabilistic sampling strategy was used to create a broadly representative sample of the population for this age group. We used the 3.0 data release.

For the main analyses, we focused on the 3814 (with grades) to 4459 (with *g* scores) individuals who were parentally identified as Black, Hispanic, Native American, or Other; we also included individuals who were marked as belonging to multiple race/ethnic categories. The choice to focus on these ethnic groups was influenced by Marira and Mitra (2013), who note that "the most rigorous research concerning the nexus of colorism and labor market outcomes has been conducted on African American and Latino populations in the United States" (*p*. 104). Although there is little ancestry-related color variability among non-Hispanic White Americans, we also ran the analyses including this group; we relegated most of these results to the supplementary file. We excluded anyone who was parentally identified as East Asian, South Asian, or Pacific Islander primarily because there were potential problems with reliable and interpretable East and South Asian ancestry estimates. Among South Asians, admixture estimates capture both recent and archaic Indo-European admixture, rendering the interpretation of these estimates unclear. Moreover, since we were unable to create a separate Pacific Islander ancestry component due to a lack of reference samples, East Asian and Pacific Islander ancestry was confounded. In contrast, the interpretation of European, African, and Amerindian admixture among Black, Hispanic, and Native American populations is unambiguous since this admixture occurred within the last 500 years, following the Age of Discovery and the settling of the New World.

2.2. Variables

A number of theoretically relevant variables were used. They are described in the sections below.

2.2.1. Admixture estimates

The ABCD Research Consortium conducted the imputing and genotyping using Illumina XX. Quality control was performed using PLINK 1.9; a total of 516,598 genetic variants survived the quality control. When computing admixture estimates, we used only directly genotyped, bi-allelic, autosomal SNP variants (494,433 before, 493,196 after lifting). We filtered variants in the reference population dataset to reduce bias from sample non-representativeness. Variants were pruned for linkage disequilibrium at the 0.1 R^2 level using PLINK 1.9 (--indep-pairwise 10000 100 0.1), leaving 99,642 variants after pruning. Next, target samples from ABCD were merged with reference population data from 1000 Genomes and HapMap. We excluded the following 1000 Genomes and HGDP reference populations: Adygei, Balochi, Bedouin, Bougainville, Brahui, Burusho, Druze, Hazara, Makrani, Mozabite, Palestinian, Papuan, San, Sindhi, Uygur, and Yakut. These populations were excluded because either they were overly admixed or because the individuals in the ABCD sample lacked significant portions of these ancestries (as in the case of Melanesians and San). The ABCD target sample was then split into 50 random subsets (of approximately 222 persons each) and merged sequentially with the reference data. Repeat subsetting was done to avoid skewing the admixture algorithm to European ancestry, as this ancestry was dominant in the ABCD sample. Next, we performed cluster analysis and estimated ancestry based on a $k = 5$ solution (European, Amerindian, African, East Asian, and South Asian ancestries), as this provided the most comprehensive yet also parsimonious model of the US population and captured all predominant ancestral backgrounds in the US population. Our European, African, and Amerindian estimates perfectly correlated with the estimates provided in the ABCD dataset (genetic_af_european, genetic_af_african, and genetic_af_european). As ABCD does not clearly document the construction of these estimates, we used our own ancestry estimates instead.

2.2.2. General cognitive ability, 11 cognitive tests, and NIHTBX fluid and crystal composite scores

The baseline ABCD data contained the following cognitive tests: Picture Vocabulary, Flanker, List Sorting, Card Sorting, Pattern Comparison, Picture Sequence Memory, Oral Reading Recognition, Wechsler Intelligence Scale for Children's Matrix Reasoning, the Little Man Test (efficiency score), the Rey Auditory Verbal Learning Test (RAVLT) immediate recall, and RAVLT delayed recall. The first seven of these are from the NIH Toolbox® cognitive battery. For details about these measures, see Thompson et al. (2019). In addition, ABCD provides a precomputed measure of crystallized cognitive ability (based on the Picture Vocabulary Test and the Oral Reading Recognition Test) and a measure of fluid cognitive ability (based on Flanker, List Sorting, Card Sorting, Pattern Comparison, and Picture Sequence Memory) measures. Details of these measures are provided by Akshoomoff et al. (2013). The crystallized ability subtests are said to be more dependent on learning experience and "represent accumulated store of verbal knowledge and skills, and thus are more heavily influenced by education and cultural exposure, particularly during childhood" (Akshoomoff, 2013, *p*. 120).

We computed *g* scores via the multi-group confirmatory factor-analytic method detailed in Fuerst et al. (2021). In this earlier publication, we outputted the *g* scores from the best-fitting and, additionally most-parsimonious model. In this case, *g* alone was found to explain the mean parentally-identified race and ethnicity (henceforth 'race and ethnicity') differences. Scores were standardized ($M = 0.00$; $SD = 1.00$) on the total sample of 10,370 children.

2.2.3. Grades

Parents were asked: "What kind of grades does your child get on average?" (1 = As /2 = Bs /3 = Cs / 4 = Ds / 5 = Fs). We recoded this variable using a 4-point scale as is commonly used in the US (with 4.0 representing an A, 3.0 representing a B, etc.), and then standardized the scores ($M = 0.00$; $SD = 1.00$). These data were available for $N = 9128$ for the Black-Hispanic-Other-White sample and $N = 3814$ for the Black-Hispanic-Other sample.

2.2.4. Child US-born and immigrant family

Parents reported if the child was born in the United States. This variable is recoded as "1" if the child was born in the United States and "0" for all other responses. Additionally, parents reported if any family members (including the child's maternal or paternal grandparents) were born outside of the United States. This variable, immigrant family, was also recoded as "1" if any family member was born outside the United States and "0" for all other responses.

2.2.5. Sex

Parents identified the sex of the children as female or male. Sex was recoded as "1" for females, and "0" for males.

2.2.6. Age

Age was calculated starting with age in months at the time of the interview ("interview age") divided by 12.

2.2.7. General socioeconomic status (SES)

Using Principal Components Analysis (PCA), we computed a general factor of SES based on seven substantially correlated indicators, which explained 42% of the variance. The loadings on the first factor were: financial adversity (.31), area deprivation index (.49), neighborhood safety protocol (.31), parental education (.54), parental income (.66), parental marital status (.43), and parental employment status (.23). We used the *PCAmixdata* R package (Chavent et al., 2014) to analyze these data, since this algorithm handles mixed categorical and continuous data. Fuerst et al. (2021) provide more details on this variable.

2.2.8. Predicted European appearance

As interviewer-rated phenotype is often unreliable and influenced by interviewer characteristics and as we only had phenotypic data for a small subset of the sample, we created genetic predictors of phenotype for analysis of the main sample. Specifically, we used tanning ability and hair pigment polyfun scores from Weissbrod et al. (2022; note that the SNP weights for these phenotypes were first published at the beginning of 2020) and skin color, hair color, and eye color probabilities calculated using the HIrisPlex-S web application (https://hirisplex.erasmusmc.nl/). To create a genetic index of European appearance, we employed PGSs developed by Weissbrod et al. (2020; 2022) for tanning propensity and hair pigment. These were created by applying genome-wide functionally-informed fine-mapping to individuals of British descent in the UK Biobank. These scores likely better estimate causal effects, thus reducing the likely adverse effects on between-population portability stemming from the impacts of linkage phase disequilibrium differences between populations (Weissbrod et al., 2020; Weissbrod et al., 2022).

The HIrisPlex-S web application, developed for use by the US Department of Justice in forensic investigations, and validated on thousands of people from around the world (Chaitanya et al., 2018; Walsh et al., 2017; Walsh et al., 2014), imputes probabilities for skin, hair, and eye color based on 41 SNPs that are functionally related to what could broadly be termed color traits (of these 36 were for skin color, 22 for hair color, and six for eye color, with overlap). HIrisPlex-S gives probabilities that a given individual occupies a level associated with the Fitzpatrick Scale skin type (i.e., Type I, scores 0–6, "palest, freckles"; Type II, scores 7–13; Type III-IV (combined), scores 14–27; Type V, scores 28–34; Type VI, scores 35–36, this being associated with "deeply pigmented dark brown to darkest brown"). We weighted the medium score of each Type (e.g., Type I = 3) by the probability of each type to create a singular color measure. For hair color, HIrisPlex-S gives probabilities of light as opposed to dark hair color. For eye color, HIrisPlex-S gives probabilities of blue, intermediate, and dark eye color. We summed the blue and intermediate colors to create the light eye color probability.

The five phenotypic predictor scores, all based on functionally-informed SNPs, overlapped due to pleiotropy, meaning that different traits are co-influenced by a common set of genes. Because of these pleiotropic relationships, we were able to use factor analysis, yielding summary scores. Specifically, we factor-analyzed the tanning ability, hair pigment, skin color, hair color, and eye color scores, using the built-in R command *factanal.* This command function fits a common factor model using maximum likelihood estimation. A single-factor model explained 81% of the variance, and the loadings were: tanning ability .96, hair pigment .98, skin color .89, hair color .88, and eye color .79. We centered and standardized these scores in the full sample of 10,370 children. The correlation matrices are shown in Table 1.

Table 1. Correlation matrices for European genetic ancestry, predicted European appearance, and genetically predicted hair, skin, and hair color scores

a. Black-Hispanic-Other sample (N =4459)

Variable	*M*	*SD*	1	2	3	4	5	6
1. European ancestry	0.44	0.28						
2. European appearance	-0.89	0.79	.84** [.83, .85]					
3. UKBB tanning ability	-0.85	0.87	.85** [.84, .86]	.95** [.95, .96]				
4.UKBB hair pigment	-0.88	0.80	.82** [.81, .83]	.99** [.99, .99]	.92** [.92, .92]			
5. HirisPlex skin color	-0.82	0.81	.67** [.66, .69]	.82** [.81, .83]	.78** [.76, .79]	.77** [.76, .79]		
6. HirisPlex hair color	0.21	0.29	.64** [.62, .66]	.84** [.83, .84]	.75** [.73, .76]	.80** [.79, .81]	.71** [.69, .72]	
7. HirisPlex eye color	0.13	0.26	.52** [.50, .54]	.71** [.69, .72]	.59** [.57, .61]	.68** [.67, .70]	.63** [.61, .65]	.73** [.71, .74]

b. Black-Hispanic-Other-White sample (N =10,370)

Variable	*M*	*SD*	1	2	3	4	5	6
1. European ancestry	0.75	0.33						
2. European Appearance	0.00	1.00	.89** [.89, .89]					
3. UKBB tanning ability	0.00	1.00	.89** [.89, .89]	.97** [.96, .97]				
4. UKBB hair pigment	0.00	1.00	.88** [.88, .89]	.99** [.99, .99]	.94** [.94, .95]			
5. HirisPlex skin color	0.00	1.00	.80** [.79, .80]	.90** [.89, .90]	.87** [.87, .88]	.87** [.86, .87]		
6. HirisPlex hair color	0.53	0.39	.77** [.76, .78]	.89** [.89, .89]	.82** [.82, .83]	.87** [.86, .87]	.77** [.77, .78]	
7. HirisPlex eye color	0.44	0.41	.66** [.65, .67]	.80** [.79, .80]	.70** [.69, .71]	.78** [.77, .79]	.74** [.73, .75]	.79** [.79, .80]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation. * indicates $p < .05$. ** indicates $p < .01$.

2.2.9. European Phenotype

The ABCD twin data contained race-related phenotypic ratings for N = 239 individuals (after removing one MZ twin from each MZ twin pair). There were two ordinal-scale ratings for each phenotype. The phenotypes were as follows: (1) Hair color ("zyg_ss_t1_hair_dark"; "zyg_ss_t2_hair_dark") which was rated as "0" = light, "1" = medium, "2" = dark, and which we reverse-coded. (2) Hair pigment ("zyg_ss_t1_hair_col"; "zyg_ss_t2_hair_col") which was rated as "0" = light blond, "1" = blond, "2" = red, "3" = brown, "4" = black, and which we reverse-coded. (3) Eye color ("zyg_ss_t1_eye_col"; "zyg_ss_t2_eye_col"); this scale was originally rated as "0" = blue, "1" = gray, "2" = green, "3" = hazel, "4" = brown, and we recoded this scale as follows: light (blue, grey, and green) = "1" and dark (hazel & brown) = "0". (4) Hair form, based on the sum of scores from hair texture ("zyg_ss_t1_hair_txtr"; "zyg_ss_t2_hair_txtr"), which was rated as "0" = coarse, "1" = medium, "2" = fine, and hair type ("zyg_ss_t1_hair_type"; "zyg_ss_t2_hair_type") which was rated as "0" = curly, and "1" = wavy, "2" = straight. Higher values were associated with a more typical European hair form.

Since we had data from two raters, we were able to compute reliability estimates. Internal consistency coefficients (ICC) estimates and their 95% confidence intervals were calculated using the *Psych* statistical package (Revelle & Revelle, 2015). These values were based on a mean-rating (k = 2), absolute-agreement, one-way random model (i.e., ICC1k; Koo & Li, 2016). The one-way model used phenotypic data sourced from two data collection sites (02 and 19). The one-way model is recommended in cases such as these (Koo & Li, 2016). Hair color, hair pigment, eye color, and hair form had average reliabilities [and confidence intervals] of .67 [.59, .73], .68 [.82, .88], 84 [.80, .87], and .89 [.86, .91], respectively. These values indicate moderate to good reliability (Koo & Li, 2016), so we used the average of ratings in subsequent analyses. Moreover, the four phenotype scores were strongly correlated with European ancestry (rs = .44 to .57); the correlation matrices are shown in Table 2.

To create rated-European phenotype scores, we factor-analyzed the scores using the R command *factanal*. A single-factor model explained 35% of the variance, and the loadings were as follows: hair color .46, hair pigment .74, eye color .71, and hair form .36. We centered and standardized these scores on the subsample of 239 individuals with phenotypic data.

The correlation between our genetically predicted European appearance and rated European phenotype was r = .76, which can be regarded as high-magnitude (Gignac & Szodorai, 2016). Notably this correlation was higher than that with global ancestry at r = .71, as would be expected if our variable predicted physical appearance above and beyond genetic ancestry. The ICCs for genetically predicted European appearance and rated European phenotype (N = 239) were .80 and .89 for single and average raters, respectively. For the Black-Hispanic-Other subsample (N = 88), these ICCs were .75 and .86, respectively. These magnitudes are usually interpreted to mean good reliability (Koo & Li, 2016).

Table 2. Correlation matrices for European genetic ancestry, genetically predicting European appearance, phenotypic European appearance, and specific phenotypes in the twin sample with phenotypic ratings ($N = 239$)

Variable	M	SD	1	2	3	4	5	6
1. European ancestry	0.85	0.21						
2. Predicted European appearance	0.00	1.00	.79**					
3. Phenotypic European appearance	0.00	1.00	.71**	.76**				
4. Phenotype: light hair	0.00	1,00	.44**	.48**	.54**			
5. Phenotype: hair pigment	0.00	1.00	.56**	.61**	.87**	.34**		
6. Phenotype: light eyes	0.00	1.00	.57**	.64**	.83**	.32**	.53**	
7. Phenotype: fine hair form	0.00	1.00	.53**	.43**	.43**	.21**	.26**	.25**

Note. M and SD are used to represent mean and standard deviation, respectively. * indicates $p < .05$. ** indicates $p < .01$. $N = 239$.

2.2.10. Race and ethnicity fraction and Hispanic

Based on the 18 questions asking about the child's race, we created four dummy race and ethnicity variables: Black, White, Native American, and Not Otherwise Classified (NOC).

The NOC category included those identified as "Other Race," "Refused to answer," or "Don't Know". Asians and Pacific Islanders were previously excluded and so were not included in the NOC category. We transformed these variables into interval race and ethnicity variables. These were calculated as the value selected for each of the four groups (0 or 1) over the total number of responses (0 to 4). Thus, individuals were assigned four race and ethnicity fractions ranging from 0 to 1. We used the White interval variable as the benchmark group. As a result, this variable is dropped from the regression models. As with ancestry, we leave these variables unstandardized so that the unstandardized beta coefficients for race and ethnicity fraction can be interpreted as a change in 100 percent race and ethnicity identity for every standardized unit of the dependent variable. We further create a variable for Hispanic ethnicity, coded as "1" for "Hispanic" and "0" for non-Hispanic. The correlation matrices for the predicted European appearance, genetic ancestry, and race and ethnicity variables, based on the Black-Hispanic-Other-White sample, are shown in Table 3.

Table 3. Correlation matrices for predicted European appearance, genetic ancestry, and race and ethnicity in the Black-Hispanic-Other-White sample ($N = 10,370$)

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Predicted European appearance	0.00	1.00								
2. European ancestry	0.75	0.33	.89**							
3. African ancestry	0.18	0.31	-.78**	-.89**						
4. Amerindian Ancestry	0.06	0.14	-.32**	-.31**	-.13**					
5. frac White	0.73	0.43	.77**	.86**	-.83**	-.13**				
6. frac Black	0.20	0.38	-.72**	-.83**	.95**	-.17**	-.82**			
7. frac Native American	0.02	0.11	-.03**	-.03**	-0.01	.08**	-.18**	-.05**		
8. frac NOC	0.06	0.23	-.21**	-.19**	-.03**	.49**	-.41**	-.13**	-.04**	
9. Hispanic	0.19	0.40	-.27**	-.22**	-.12**	.76**	-.07**	-.17**	.04**	.38**

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. * indicates $p < .05$. ** indicates $p < .01$.

2.2.11. eduPGS

To create educational polygenic scores (PGS), we scored the genomes using PLINK v1.90b6.8. We used the multi-trait analysis of genome-wide association study (MTAG) eduPGS SNPs ($N = 8,898$ variants in this sample) to compute eduPGS, and we based ourselves on the results from the genome-wide association study (GWAS) of Lee et al. (2018). These scores were based on educational attainment ($N = 1,131,881$), cognitive ability ($N = 257,841$), hardest math class taken ($N = 430,445$), and mathematical ability ($N = 564,698$) (Lee et al., 2018). Previous research has shown these scores to have good predictive validity in European populations, and reasonable trans-ethnic predictive validity in Hispanic, and African-American populations (Fuerst et al., 2021; Lasker et al., 2019).

2.3. Analyses

2.3.1. Validation of European Phenotypic Predictor

To validate our genetic predictor of European appearance, we ran two analyses. First, we ran regression models in which we predicted parentally-identified-child-racial category (White = "1"; non-White = "0") based on our European phenotypic predictor, with European genetic ancestry controlled. While racial classifications in the US are primarily based on perceived continental ancestry (Harris, 2008; Hall, 2020), skin color has been found to influence the classifications independently of ancestry (Schachter, Flores & Maghbouleh, 2021). Thus, we expected European appearance to predict the identified race of the child, independently of genetic ancestry.

In line with the recommendation of Heeringa and Berglund (2020), we used a linear mixed-effects model rather than ordinary least squares. Linear mixed-effects involve partially decomposing the residual term into linear random effects components linked to the data-collection-site identifiers and same-family identifiers within the sample. This allows for the possibility of error term correlations within data collection sites or within families with multiple tested individuals (see:

Heeringa & Berglund, 2020). This specification replicates that which was used by the ABCD Data Exploration and Analysis Portal (DEAP), and so the use of this multilevel model also aids in replication. To run these analyses, we employed the *lmer* command from the **lme4** package (Bates, Mächler, Bolker, & Walker, 2015).

As noted in Section 2.2.9 our European phenotype predictor had excellent reliability. So, second, using the subsample for which we had phenotypic ratings, we examine the degree to which our genetically predicted European appearance variable predicted rated European phenotype independently of genetic ancestry and, in supplemental analyses, race and ethnicity. We expect our appearance predictor to have validity as a predictor of rater reported appearance above and beyond global genetic ancestry. Since, we only had two sample sites and since we dropped one of two MZ twins we ran ordinary least square (OLS) for simplicity. However, we also report results based on linear mixed-effects, controlling for site, in Tab 2 of the supplementary file.

2.3.2. Main regression analyses

We next ran a series of regression analyses, using mixed-effects models, in which *g* or parentally-reported grades was the dependent variable and European appearance was the main independent variable. We ran five models, which sequentially added (parentally-identified) racial and ethnic category, SES, and ancestry, and both ancestry and SES to the initial regression model. The expectation was that if the colorism model was correct, predicted European appearance would retain validity when we added race and ethnicity, SES, and ancestry. In running these models, we followed a methodology similar to that of Telles and Paschel (2014) and Telles, Flores, and Urrea-Giraldo (2015), except that we also added genetic ancestry to the models. So, we followed established models of analyzing these kinds of data.

For the analyses with *g* (and also cognitive subtests), we used the mixed-effects model discussed above. However, parent-reported school grades showed a strong censoring effect (A = 49%; B = 36%; C to F = 15%), with many individuals receiving the highest grade of an A and the distribution not being normal. Therefore, we also ran analyses for grades using tobit regression, which adjusts for censored data, since censoring can induce effects resembling model misspecification. For these analyses, we used the *tobit* function in the **AER** package (Kleiber et al., 2020). No package for multilevel tobit regression is available for R, so for these tobit regression analyses we dropped the site and family random effects and included dummy variables for recruitment site instead. This is a suboptimal method, so we ran all analyses using both the standard mixed-effects model based on linear regression and the tobit regression models and compared the results. We also assessed the degree of collinearity for the European appearance variable, using the variance inflation factor (VIF) statistic. We used the **car** statistical package (Fox and Weisberg, 2011) to calculate the values of VIF.

2.3.3. Causal mediation analyses

Following the suggestion of Huddleston and Montgomery (2010), we ran causal mediation analysis using the **mediation** R package (Tingley et al., 2013). This package estimates both the mediation effect and the proportion of the effect that is directly causally mediated. We ran analyses with *g* as a mediator of the relation between grades and either European ancestry or European appearance. We ran these analyses both in the Black-Hispanic-Other-White sample and the Black-Hispanic-Other sample. European ancestry, European appearance, sex, age, child US-born, immigrant family, race and ethnicity, and interview site are included as covariates in all of the mediation analyses. Since this package cannot handle two levels of random effects, we dropped the family level from the analyses when using the mixed-effects model. We alternatively ran the models using linear regression for *g* and tobit regression for grades, in which case site effects were included as dummy variables.

2.3.4. Robustness analyses

Finally, as a robustness check, we reran the main regression analyses using the UKBB tanning and the HIrisPlex skin color variables instead of European appearance. This was done because theorists of colorism put the causal role of skin-color-based discrimination central (Marira & Mitra, 2013), and it may be that our European appearance variable is obscuring the relation between skin color and academic outcomes by also taking into account hair and eye color. We also reran the main analyses on the Hispanic and Black subsamples independently to determine if effects were present in subgroups. This was done because researchers using sibling designs have reported significant appearance-related effects among Hispanics (Ryabov, 2016) without these effects being present in the full sample (Kizer, 2017). It may be that there are appearance-related effects among Hispanics or Blacks but not in the combined sample.

Additionally, we ran the analyses on all eleven cognitive subtests used to create *g* scores, and the NIH toolbox crystallized and fluid ability scales since it is possible that discriminatory effects will be localized on certain measures of cognitive ability. Differences might be expected to be more pronounced on crystallized measures since crystallized intelligence is “more heavily influenced by education and cultural exposure” (Akshoomoff, 2013, *p*. 120) and since theorists of colorism argue that colorism will manifest as a barrier to learning opportunities.

We further tested for possible cross-trait assortative mating by examining if eduPGS scores predict European appearance, tanning, and color independent of European ancestry. If there was cross-trait assortative mating, we would expect a substantial effect of eduPGS on these traits independent of European genetic ancestry. Norris (2019) reports polygenic evidence of assortative mating for education among Latin-American populations, while Zou et al. (2015) report evidence of assortative mating on European appearance. So, it is plausible that there was simultaneous assortative mating on both European appearance and education. While generally not mentioned by theorists of colorism, cross-trait assortative mating has long been hypothesized to explain advantages associated with European appearance in admixed American populations (e.g., Jensen, 1998; Reuter, 1917; Valenzuela, 2011).

2.4. Data and code

The R-code and additional model outputs are included in the supplementary file, available at Open Science Frame: https://osf.io/jqkns/. ABCD data are available to qualified researchers at: https://nda.nih.gov/abcd.

3. Results

3.1. Validation

Table 4 shows the regression results for predicting White racial and ethnic categorization. Model 1 shows the results for the Black-Hispanic-Other sample, while Model 2 shows the results for the

Black-Hispanic-Other-White sample. Model 1a and Model 2a include only European ancestry as the predictor, while Model 1b and Model 2b include all non-European ancestries as predictors. We note that our genetically-predicted European appearance variable predicts participant race and ethnicity independently of genetic ancestry. The effect is small, as expected, because race and ethnicity in the US are mostly understood in terms of ancestry.

Table 4. Regression models with genetically-predicted European appearance as the key independent variable and white racial and ethnic categorization as the dependent variable using the Black-Hispanic-Other sample (Model 1) and the Black-Hispanic-Other-White sample (Model 2)

	Model 1a: Race/ethnicity, Appearance, & European Ancestry		**Model 1b: Race/ethnicity, Appearance, & non-European Ancestries**		**Model 2a: Race/ethnicity, Appearance, & European Ancestry**		**Model 2b: Race/ethnicity, Appearance, & non-European Ancestries**	
Predictors	B	*p*	B	*p*	B	*p*	B	*p*
(Intercept)	-0.10 (0.04)	**0.005**	0.76 (0.03)	**<0.001**	-0.14 (0.02)	**<0.001**	0.94 (0.02)	**<0.001**
Predicted European appearance	0.03 (0.01)	**0.004**	0.02 (0.01)	**0.017**	0.02 (0.00)	**<0.001**	0.02 (0.00)	**<0.001**
European ancestry	0.88 (0.03)	**<0.001**			1.07 (0.01)	**<0.001**		
Child US Born	0.04 (0.02)	0.071	0.04 (0.02)	0.124	0.05 (0.01)	**0.001**	0.04 (0.01)	**0.003**
Immigrant Family	0.14 (0.01)	**<0.001**	0.06 (0.01)	**<0.001**	0.07 (0.01)	**<0.001**	0.02 (0.01)	**<0.001**
African ancestry			-0.92 (0.03)	**<0.001**			-1.13 (0.01)	**<0.001**
Amerindian ancestry			-0.35 (0.04)	**<0.001**			-0.67 (0.02)	**<0.001**
East Asian ancestry			-0.96 (0.13)	**<0.001**			-0.81 (0.08)	**<0.001**
South Asian ancestry			-1.14 (0.21)	**<0.001**			-1.00 (0.13)	**<0.001**
Random Effects								
σ^2	0.02		0.01		0.01		0.01	
τ_{00}	0.08 site_id_l:rel_family_id 0.00 site_id_l		0.07 site_id_l:rel_family_id 0.01 site_id_l		0.04 site_id_l:rel_family_id 0.00 site_id_l		0.03 site_id_l:rel_family_id 0.00 site_id_l	
ICC	0.85		0.84		0.8		0.79	
N	22 site_id_l 3863 rel_family_id		22 site_id_l 3863 rel_family_id		22 site_id_l 8672 rel_family_id		22 site_id_l 8672 rel_family_id	
Observations	4459		4459		10370		10370	
Marginal R^2 / Conditional R^2	0.464 / 0.918		0.527 / 0.925		0.739 / 0.948		0.754 / 0.948	

Note: Unstandardized beta coefficients (B) and *p*-values (*p*) from the mixed-effects models, with recruitment site and family common factors treated as random effects are shown. The values in parentheses are standard errors. The marginal and conditional R^2s of the mixed-effects model are shown at the bottom. ICC = Intraclass Correlation Coefficient.

Next, Table 5 shows the regression results for predicting interviewer-rated phenotype based on genetically predicted European appearance in the small twin sample that had phenotypic ratings. Model 1 shows the results without genetic ancestry, while models 2a and 2b add European ancestry and non-European ancestries, respectively. We note that our genetically predicted European appearance variable predicts interviewer-rated European phenotype over and above genetic ancestry. The results, for the equivalent Model 2a, are substantially the same when we subset to the 88 Blacks, Hispanics, and Others in the twin sample (Predicted European appearance B = .59; *S.E.* = .11) or when we additionally include parental reported race in the models. These additional results are provided in Table 2 of the supplementary file.

Table 5. Regression models with genetically predicted European appearance as the key independent variable and interviewer-rated phenotype as the dependent variable

	Model 1: Phenotype ~ Predicted Phenotype		**Model 2a: Phenotype ~ Predicted Phenotype w/European ancestry**		**Models 2b: Phenotype ~ Predicted Phenotype w/non-European Ancestries**	
Predictors	B	*p*	B	*p*	B	*p*
(Intercept)	0.04 (0.06)	0.469	-1.09 (0.27)	**<0.001**	0.28 (0.09)	**0.001**
European appearance	0.76 (0.04)	**<0.001**	0.54 (0.07)	**<0.001**	0.53 (0.07)	**<0.001**
Sex	-0.09 (0.08)	0.286	-0.08 (0.08)	0.309	-0.07 (0.08)	0.360
European ancestry			1.32 (0.31)	**<0.001**		
African ancestry					-1.14 (0.33)	**0.001**
Amerindian ancestry					-1.61 (0.45)	**<0.001**
East Asian ancestry					-2.18 (3.10)	0.483
South Asian ancestry					-11.22 (13.97)	0.423
Observations	239		239		239	
R^2 / R^2 adjusted	0.584 / 0.581		0.614 / 0.610		0.619 / 0.610	

Notes: Unstandardized beta coefficients (B) and *p*-values (*p*) from the OLS regression models. The values in parentheses are standard errors. The marginal and conditional R^2 of the mixed effects model are shown at the bottom.

3.2. Main results

Table 6 shows the main results with *g* as the dependent variable for the Black-Hispanic-Other sample. Its Model 1 reveals that the relation between predicted European appearance and *g* is statistically significant. When we add race and ethnicity in Model 2, this relation is reduced but still

Table 6. Regression results for the effect of predicted European appearance on *g* in the Black-Hispanic-Other sample

	M1: *g* ~ European appearance		**M2: *g* ~ European appearance + race and ethnicity**		**M3: *g* ~ European appearance + race and ethnicity + SES**		**M4a: *g* ~ European appearance + race and ethnicity + European ancestry**		**M4b: *g* ~ European appearance + race and ethnicity + European ancestry + SES**	
Predictors	B	*p*	B	*p*	B	*p*	B	*p*	B	*p*
Intercept	0.31 (0.25)	0.22	0.42 (0.26)	0.11	0.38 (0.25)	0.14	-0.62 (0.28)	**0.03**	-0.28 (0.27)	0.30
Age	-0.05 (0.02)	**0.03**	-0.05 (0.02)	**0.03**	-0.05 (0.02)	**0.03**	-0.04 (0.02)	0.06	-0.05 (0.02)	**0.04**
Sex	0.02 (0.03)	0.45	0.03 (0.03)	0.37	0.04 (0.03)	0.22	0.02 (0.03)	0.44	0.03 (0.03)	0.26
Predicted European appearance	0.34 (0.02)	**<0.01**	0.25 (0.03)	**<0.01**	0.15 (0.03)	**<0.01**	0.03 (0.04)	0.41	0.01 (0.03)	0.72
Child US Born	0.05 (0.08)	0.55	0.05 (0.08)	0.54	0.13 (0.08)	0.09	0.07 (0.08)	0.37	0.14 (0.08)	0.07
Immigrant Family	0.25 (0.04)	**<0.01**	0.20 (0.04)	**<0.01**	0.12 (0.04)	**0.04**	0.22 (0.04)	**<0.01**	0.14 (0.04)	**0.01**
Frac Black			-0.31 (0.07)	**<0.01**	-0.23 (0.07)	**0.02**	0.08 (0.08)	0.36	0.02 (0.08)	0.83
Frac Native American			-0.10 (0.12)	0.40	-0.05 (0.11)	0.66	-0.07 (0.12)	0.58	-0.03 (0.11)	0.78
Frac NOC			-0.19 (0.06)	**0.01**	-0.07 (0.05)	0.18	-0.07 (0.06)	0.23	-0.00 (0.05)	0.93
Hispanic			-0.05 (0.06)	0.38	-0.01 (0.05)	0.86	-0.04 (0.06)	0.45	-0.01 (0.05)	0.86
SES					0.35 (0.02)	**<0.01**			0.32 (0.02)	**<0.01**
European ancestry							1.24 (0.13)	**<0.01**	0.79 (0.12)	**<0.01**

Random Effects					
σ^2	0.52	0.52	0.51	0.51	0.51
τ_{00}	0.52 site_id_l:rel_family_id	0.52 site_id_l:rel_family_id	0.43 site_id_l:rel_family_id	0.50 site_id_l:rel_family_id	0.43 site_id_l:rel_family_id
	0.02 site_id_l	0.02 site_id_l	0.04 site_id_l	0.02 site_id_l	0.04 site_id_l
ICC	0.51	0.51	0.48	0.51	0.48
N	22 site_id_l	22 site_id_l	22 site_id_l	22 site_id_l	22 site_id_l
	3863 rel_family_id	3863 rel_family_id	3863 rel_family_id	3863 rel_family_id	3863 rel_family_id
Observations	4459	4459	4459	4459	4459
Marginal R^2 / Conditional R^2	0.085 / 0.549	0.095 / 0.556	0.180 / 0.571	0.116 / 0.565	0.187 / 0.574

Notes: Unstandardized beta coefficients (B) and *p*-values (*p*) from the mixed-effects models, with recruitment site and family common factors treated as random effects are shown. The values in parentheses are standard errors. The marginal and conditional R^2 of the mixed effects model are shown at the bottom. ICC = Intraclass Correlation Coefficient.

statistically significant. When we add SES, in Model 3, the effect is further reduced but still statistically significant. However, when we add genetic ancestry in Model 4a (without SES) and 4b (with SES) the relation becomes statistically non-significant. The results for the Black-Hispanic-Other-White sample, provided in the supplementary material, also reveal no statistically significant effect when adding European ancestry in Model 4a (without SES) and 4b (with SES).

Table 7 shows the main results, again for the Black-Hispanic-Other sample, using grades as the dependent variable. In this table, tobit regression is used. Model 1 reveals that the relation between predicted European appearance and grades is statistically significant. When we add race and ethnicity in Model 2, this relation is reduced but is still statistically significant. When we add SES, in Model 3, the effect is further reduced but is still statistically significant. Adding genetic ancestry in Models 4a and 4b (with SES) does not change this.

However, in the final Model (4b), the relation is only marginally statistically significant (p = 0.042). In contrast, the results for the Black-Hispanic-Other-White sample, provided in the supplementary material, show no statistically-significant effect of European appearance in Model 4a (without SES) and 4b (with SES). Moreover, the effects of European appearance in the Black-Hispanic-Other-White sample are trivial. Thus, the effect of European appearance on parent-reported grades only shows up in the Black-Hispanic-Other sample.

Table 7. Tobit regression results for the effect of predicted European appearance on grades in the Black-Hispanic-Other sample

	M1: Grades ~ European appearance		**M2: Grades ~ European appearance + SIRE**		**M3: Grades ~ European appearance + SIRE + SES**		**M4a: Grades ~ European appearance + SIRE + European ancestry**		**M4b: Grades ~ European appearance + SIRE + European ancestry + SES**	
Predictors	B	*p*	B	*p*	B	*p*	B	*p*	B	*p*
Age	-0.03 (0.03)	0.27	-0.03 (0.03)	0.30	-0.04 (0.03)	0.24	-0.03 (0.03)	0.34	-0.03 (0.03)	0.25
Sex	0.29 (0.04)	**<0.01**	0.29 (0.04)	**<0.01**	0.29 (0.04)	**<0.01**	0.29 (0.04)	**<0.01**	0.29 (0.04)	**<0.01**
European appearance	0.21 (0.02)	**<0.01**	0.20 (0.03)	**<0.01**	0.12 (0.03)	**<0.01**	0.10 (0.04)	**0.01**	0.09 (0.04)	**0.04**
Child US Born	-0.07 (0.10)	0.50	-0.08 (0.10)	0.43	-0.03 (0.10)	0.77	-0.07 (0.10)	0.47	-0.03 (0.10)	0.78
Immigrant Family	0.07 (0.04)	0.12	0.08 (0.05)	0.12	0.03 (0.05)	0.55	0.09 (0.05)	0.08	0.03 (0.05)	0.49
Frac Black SIRE			0.00 (0.09)	0.99	0.05 (0.09)	0.54	0.16 (0.10)	0.11	0.11 (0.10)	0.25
Frac Native American SIRE			0.12 (0.14)	0.36	0.15 (0.13)	0.27	0.13 (0.14)	0.33	0.15 (0.13)	0.26
Frac NOC SIRE			-0.13 (0.07)	0.06	-0.06 (0.07)	0.40	-0.08 (0.07)	0.23	-0.04 (0.07)	0.55
Hispanic			0.03 (0.07)	0.62	0.07 (0.07)	0.27	0.04 (0.07)	0.55	0.07 (0.07)	0.26
SES					0.22 (0.02)	**<0.01**			0.22 (0.02)	**<0.01**
European ancestry							0.50 (0.15)	**0.01**	0.19 (0.15)	0.20
Observations	3814		3814		3814		3814		3814	
R^2 Nagelkerke	0.056		0.058		0.089		0.061		0.089	

Notes: Unstandardized beta coefficients (B) and *p*-values (*p*) from the tobit models, with recruitment site (not shown) added as dummy variables. The values in parentheses are standard errors.

Table 8, Model 1 shows the alternative multi-level regression results. For these, when we add genetic ancestry in Models 4a and 4b (with SES), the relation becomes statistically non-significant, although the model approaches statistical significance in the case of 4a (without SES) (*p* = 0.06).

Table 8. Mixed-effects regression results for the effect of predicted European appearance on grades in the Black-Hispanic-Other Sample

	M1: Grades ~ European appearance		**M2: Grades ~ European appearance + race and ethnicity**		**M3: Grades ~ European appearance + race and ethnicity + SES**		**M4a: Grades ~ European appearance + race and ethnicity + European ancestry**		**M4b: Grades ~ European appearance + race and ethnicity + European ancestry + SES**	
Predictors	B	*p*	B	*p*	B	*p*	B	*p*	B	*p*
(Intercept)	0.45 (0.29)	0.11	0.42 (0.29)	0.15	0.46 (0.29)	0.11	-0.03 (0.31)	0.93	0.26 (0.31)	0.40
Age	-0.06 (0.03)	**0.03**	-0.06 (0.03)	**0.04**	-0.06 (0.03)	**0.02**	-0.05 (0.03)	**0.04**	-0.06 (0.03)	**0.02**
Sex	0.34 (0.03)	**<0.01**	0.34 (0.03)	**<0.01**	0.34 (0.03)	**<0.01**	0.34 (0.03)	**<0.01**	0.34 (0.03)	**<0.01**
Predicted European appearance	0.19 (0.02)	**<0.01**	0.17 (0.03)	**<0.01**	0.10 (0.03)	**0.01**	0.08 (0.04)	0.06	0.06 (0.04)	0.13
Child US Born	-0.15 (0.09)	0.10	-0.17 (0.09)	0.07	-0.11 (0.09)	0.24	-0.15 (0.09)	0.09	-0.10 (0.09)	0.26
Immigrant Family	0.04 (0.04)	0.36	0.04 (0.05)	0.38	-0.02 (0.05)	0.62	0.05 (0.05)	0.27	-0.02 (0.05)	0.73
Frac Black			-0.00 (0.08)	0.97	0.05 (0.08)	0.54	0.16 (0.09)	0.08	0.12 (0.09)	0.19
Frac Native American			0.03 (0.13)	0.81	0.07 (0.12)	0.60	0.04 (0.13)	0.74	0.07 (0.12)	0.58
Frac NOC			-0.20 (0.06)	**0.01**	-0.13 (0.06)	**0.03**	-0.14 (0.06)	**0.02**	-0.11 (0.06)	0.07
Hispanic			0.05 (0.06)	0.40	0.07 (0.06)	0.26	0.06 (0.06)	0.34	0.07 (0.06)	0.24
SES					0.22 (0.02)	**<0.01**			0.21 (0.02)	**<0.01**
European_ancestry							0.53 (0.14)	**<0.01**	0.23 (0.14)	0.10

Random Effects					
σ^2	0.76	0.75	0.75	0.75	0.75
τ_{00}	0.32 site_id_l:rel_family_id	0.32 site_id_l:rel_family_id	0.29 site_id_l:rel_family_id	0.32 site_id_l:rel_family_id	0.29 site_id_l:rel_family_id
	0.01 site_id_l	0.01 site_id_l	0.00 site_id_l	0.01 site_id_l	0.00 site_id_l
ICC	0.30	0.30	0.28	0.30	0.28
N	22 site_id_l	22 site_id_l	22 site_id_l	22 site_id_l	22 site_id_l
	3296 rel_family_id	3296 rel_family_id	3296 rel_family_id	3296 rel_family_id	3296 rel_family_id
Observations	3814	3814	3814	3814	3814
Marginal R^2 / Conditional R^2	0.048 / 0.338	0.051 / 0.339	0.086 / 0.339	0.056 / 0.343	0.087 / 0.340

Notes: Unstandardized beta coefficients (B) and *p*-values (*p*) from the mixed-effects models, with recruitment site and family common factors treated as random effects, are shown. The values in parentheses are standard errors. The marginal and conditional R^2s of the mixed-effects model are shown at the bottom. ICC = Intraclass Correlation Coefficient.

The results for the Black-Hispanic-Other-White sample, provided in the supplementary material, show no statistically significant effect of European appearance in Model 4a (without SES) and 4b (with SES). The discrepancy between the tobit and multilevel regression results could be due to not including random effects in the former or due to censoring-related bias in the latter. Regardless, the betas for both 4b models are fairly similar (B = .09 vs. B = .06) and suggest some unaccounted-for European appearance-related effect on parent-reported student grades.

We additionally computed the values of variance inflation factors (VIFs), which measure the amount of multicollinearity in a regression analysis. In cases of high multicollinearity, it can be difficult to distinguish between the effects of individual independent variables owing to high correlations among these. The full results for these tests for excessive multicollinearity are provided in Tab 5 of the supplementary file. For the Black-Hispanic-Other sample, the values of VIFs for predicted European appearance in Models 4a to 4b of Tables 6-8 ranged from 2.9 to 3.6. For the Black-Hispanic-Other-White sample, VIFs for predicted European appearance ranged from 4.0 to 4.6. Usually, a VIF value > 5 or, less strictly, > 10 is said to be problematic (James et al., 2013), but the values found here for European appearance are lower than these commonly-used thresholds, indicating that multicollinearity is not likely to be confounding our results. We also ran the models dropping the frac_SIRE variables, which, in the presence of European appearance, were leading to collinearity with European ancestry. As expected, doing so decreased the values of VIFs for European ancestry but did not substantially change the results.

3.3. Mediation analysis

The results from the causal mediation analyses are summarized in Table 9. The full results are provided in Tab 6 of the supplementary file. In both the Black-Hispanic-Other and the Black-Hispanic-Other-White samples, *g* was a substantial and statistically significant mediator of the relation between genetic ancestry and grades. In contrast, *g* was not a statistically significant mediator of the relation between European appearance and grades. In the Black-Hispanic-Other sample, European appearance had a significant total effect on grades (p = .03) with the same magnitude of effect as in Model M4a of table 7. Nonetheless, *g* was not a statistically significant mediator of the relation between predicted European appearance and grades.

Table 9. Summary of the causal mediation results for the full sample

Method: LMER

Predictor	Mediator	Criterion	Mediation Effect	Proportion of total effect mediated	N
Black-Hispanic-Other					
Predicted European appearance	*g*	Grades	0.02 [-0.01, 0.05]	0.26	4459
European Ancestry	*g*	Grades	**0.53** [0.41, 0.66]	1.03	3814
Black-Hispanic-Other-White					
Predicted European appearance	*g*	Grades	0.01 [-0.01, 0.03]	0.25	10370
European Ancestry	*g*	Grades	**0.61** [0.53, 0.71]	0.75	9128

Method: LM-Tobit

Predictor	Mediator	Criterion	Mediation Effect	Proportion of total effect mediated	N
Black-Hispanic-Other					
Predicted European appearance	*g*	Grades	0.02 [-0.01, 0.06]	0.24	4459
European Ancestry	*g*	Grades	**0.54** [0.43, 0.67]	1.12	3814
Black-Hispanic-Other-White					
Predicted European appearance	*g*	Grades	0.01 [-0.01, 0.02]	0.25	10370
European Ancestry	*g*	Grades	**0.52** [0.44, 0.60]	0.83	9128

Notes: Statistically significant results ($p < 0.05$) are presented in bold.

3.4. Robustness analyses

As a robustness check, we reran the main regression analyses using the UKBB tanning and the HIrisPlex skin color variables instead of European appearance. These results are reported in Tab 8 and Tab 9 of the supplementary file. Neither the effects of UKBB tanning nor the effects of HIrisPlex skin color were close to significant for *g* or grades. So, using a composite indicator of hair, eye, and skin tanning/color instead of skin tanning/color alone did not decrease the effects. Rather, effects, particularly for grades, were stronger when using the European appearance predictor.

Next, we reran the main analyses on both the Hispanic (N = 2021 to 1741) and non-Hispanic Black (N = 1690 to 1432) subsamples. These results are shown in Tab 10 of the supplementary file. Again, European appearance, tanning, and color were not significantly related to either *g* or grades in the Hispanic and Black subsamples.

Additionally, we ran the analyses on all eleven cognitive subtests used to create *g* scores in addition to the NIHTBX fluid and crystallized composite scores. These results, for the Black-Hispanic-Other sample, are shown in Tab 13 of the supplementary file. We found statistically significant results only for the Picture Vocabulary test (B = 0.097; *S.E.* = .033), but not the other crystallized test, namely Oral Reading Recognition (B = 0.000; *S.E.* = .034). We also found marginally significant results for the crystallized composite scores (B = 0.068; *S.E.* = .033), which were due to the highly significant results for Picture Vocabulary, since the crystallized composite was derived from the Picture Vocabulary and Oral Reading Recognition tests. These results for crystallized composite scores and for Picture Vocabulary scores could represent a real effect, since it seems more likely that appearance-based discrimination would be present on a measure of knowledge and crystallized cognitive ability than fluid ability. Alternatively, this could represent a coincidence due to multiple testing (i.e., looking at the effects on eleven independent measures).

Finally, we tested for possible cross-trait assortative mating by examining if eduPGS scores predicted European appearance independent of European ancestry. These results are shown in Tab 11 of the supplementary file. We did not find significant associations between eduPGS and European appearance independent of European ancestry, which is inconsistent with the cross-trait assortative mating hypothesis.

4. Conclusion

4.1. Discussion

The colorism model states that there is ongoing discrimination based on physical appearance in the Americas, which contributes to the link between European phenotype and better social outcomes. This model predicts that European appearance, regardless of ancestry, will have a positive impact on academic outcomes. We tested the colorism model using the admixture-regression methodology. Specifically, we hypothesized, first, that European appearance would be related to *g* when taking into account genetic ancestry, and second, that European appearance would have a relationship with grades independent of genetic ancestry. However, we found no substantial evidence to support the first hypothesis and only limited evidence to support the second. As a result, our findings do not strongly support the colorism hypothesis.

Additionally, due to meta-analyses showing a significant connection between *g* and grades, we formulated two more hypotheses: third, that *g* would play a role in the relationship between European ancestry and grades, and fourth, that *g* would play a role in the relationship between European appearance and grades. Our findings provided strong evidence for the third hypothesis, but not for the fourth hypothesis. This suggests that the relationship between European ancestry and grades is dependent on *g*, while the relationship between European appearance and grades is not dependent on *g*.

We conducted several robustness tests. When we used tanning or color, instead of European appearance, as the main predictor, we did not find any significant statistical connections between these two variables and either *g* or grades. Moreover, we did not find statistically significant associations when subsetting to the two largest admixed groups, Hispanics and Blacks. The lack of an association between either tanning or color and academic outcomes further weakens a colorism model, since this model places primacy on "the causal role of skin tone in engendering the colorism phenomenon" (Marira & Mitra, 2013, *p*. 103). So, we conclude that the outcomes are robust, and that the results of the robustness tests did not support a colorism model.

Across 15 outcome measures (*g*, grades, eleven subtests, the fluid ability composite, and the crystallized ability composite), three traits (European appearance, UKBB tanning, and HIrisPlex skin color), and four groupings (Black-Hispanic-Other-White, Black-Hispanic-Other, Black, and Hispanic) we found three statistically significant results, two of which were redundant (i.e., the effect on crystallized intelligence resulted from the effect on Picture Vocabulary). Given that 15 different overlapping academic outcomes were examined across three traits, these results could simply be chance results due to multiple testing. Moreover, the interpretation of the results for grades is complicated since the grades were parent-reported, not actual school-reported grades, and since the parent-reported grades were very course (e.g., mostly As, mostly Bs, mostly Cs, etc.).

On the other hand, the three statistically significant results may make theoretical sense. Theorists of colorism have argued that colorism will manifest as a barrier to learning opportunities (Thompson & McDonald, 2016) and optimal education (Crutchfield et al., 2022), acting against less European-appearing students. On this basis, it is reasonable to hypothesize that the effects would be larger on crystallized abilities and outcomes dependent on teachers' judgment. Therefore, future research should attempt to evaluate whether there are robust European-appearance effects on measures of crystallized ability and grades.

Regarding the general lack of association with respect to cognitive ability, one could perhaps argue that global genetic ancestry is a better predictor of overall racial appearance. Accordingly, discrimination might be based on an overall assessment of morphological differences, not just conspicuous race-associated differences in skin, eye, and hair color. However, this is somewhat different from what "colorists" have narrowly hypothesized, hence the term "colorism". If race-associated discrimination is said to be based on ancestry-indexing phenotypes, then perhaps a good alternative term might be "ancestryism". It is not clear, though, why discrimination would be finely tuned to overall phenotypic markers of genetic ancestry instead of features that are stereotypic of racial groups.

A more likely explanation for these findings is that color is associated with cognitive ability predominantly because the color phenotype proxies global ancestry. Global ancestry could be related to how differences in these characteristics are inherited or passed on from parents to children along genealogical lines. Several researchers have argued that genetic ancestry might index social disparities intergenerationally transmitted for cultural, genetic, or epigenetic reasons (Corach & Caputo, 2022; da Silva et al., 2020; Fuerst & Kirkegaard, 2016). This possibility has largely been not discussed by sociologists who have focused instead on discriminatory or diffuse cultural models concerning disparities by caste, nation, ethnicity, or race. Our results suggest that researchers should also focus on genetic ancestry and identifying the specific genetic, epigenetic, and cultural factors which mediate the intergenerational transmission of academic outcomes.

4.2. Limitations

One could argue that our genetic predictor of European appearance is not perfectly reliable. However, frequently used skin tone scales, such as the Massey and Martin (2003) scale, have been found to have low reliabilities (Hannon & DeFina, 2016; 2020), so we would argue, based on our own results, that our genetically-predicted European appearance scores appear to be more reliable than the sorts of color measures which are typically used. A genetic-based predictor of color captures phenotype over an individual's lifetime without the variability due to the low reliability of interviewer rating skin tone scales which have been found to be influenced by interviewer-rated characteristics such as race/ethnicity (Campbell et al., 2020; Cernat et al., 2019) or the interviewer's perceptions about the participants socioeconomic status (Roth et al., 2022). So, we conclude that a genetic predictor of European appearance is preferable to interviewer-rated skin tone.

It is possible that there are substantial ancestry-independent associations between color and academic outcomes in other populations or for other academic traits. The novel admixture-regression method described here can be used to investigate if this is the case.

4.3. Implications

We focused on Hispanic, Black, Other, and White Americans. This is because, according to Marira and Mitra (2013, p. 104), "the most rigorous research concerning the nexus of colorism and labor market outcomes has been conducted on African American and Latino populations in the United States". Given that the best support for colorism is said to come from the study of Black and Hispanic Americans, we would have expected the association between European appearance/color and academic outcomes to be robust in this sample. However, we did not find this to be the case, which weakens the colorism hypothesis.

4.4. Conclusions

We found little support for the colorism hypothesis. Our results suggest that the relationship between racial appearance and academic outcomes cannot be considered as prima facie evidence of discrimination, as genetic ancestry, and the inherited disadvantage associated with it, is a plausible confounding variable. In future studies on colorism, it is recommended to consider genetic ancestry as a factor in the analyses. Because discrimination is such an important topic, it is essential that more studies are carried out before we are able to draw strong conclusions.

Author contributions: All analyses were conducted by JGRF under the supervision of B. J. Pesta while at Cleveland State University (2020-2021). VS helped write and edit the paper.

Acknowledgments: Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data, but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the research results and interpretations of the authors alone and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Version 3.01. The raw data are available at https://nda.nih.gov/edit_collection.html?id=2573. Additional support for this work was made possible from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680), and Children and Screens: Institute of Digital Media and Child Development Inc. NDA waved the requirement of creating a NDA data project (NDA, Oct 26, 2022).

CRediT author statement: V. Shibaev and J. Fuerst conceived of the idea. John Fuerst performed the computations. V. Shibaev verified the analytical methods. J. Fuerst and V. Shibaev edited and revised the manuscript. Both authors discussed the results and contributed to the final manuscript.

References

Abascal, M., & Garcia, D. (2022). Pathways to Skin Color Stratification: The Role of Inherited (Dis) Advantage and Skin Color Discrimination in Labor Markets. *Sociological Science*, *9*, 346-373.

Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., ... & Heaton, R. K. (2013). VIII. NIH Toolbox Cognition Battery (CB): Composite scores of crystallized, fluid, and overall cognition. *Monographs of the Society for Research in Child Development*, *78*(4), 119-132.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bucca, M. E. (2018). *Socioeconomic Inequality Among Racial Groups in the United States* (Doctoral dissertation, Cornell University).

Campbell, M. E., Keith, V. M., Gonlin, V., & Carter-Sowell, A. R. (2020). Is a picture worth a thousand words? An experiment comparing observer-based skin tone measures. *Race and Social Problems*, *12*(3), 266-278.

Campos-Vazquez, R. M., & Medina-Cortina, E. M. (2019). Skin color and social mobility: Evidence from Mexico. *Demography*, *56*(1), 321-343.

Cernat, A., Sakshaug, J. W., & Castillo, J. (2019). The Impact of interviewer effects on skin color assessment in a cross-national context. *International Journal of Public Opinion Research*, *31*(4), 779-793.

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., ... & Walsh, S. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International: Genetics*, *35*, 123-135.

Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J. (2014). Multivariate analysis of mixed data: The R Package PCAmixdata. arXiv.1411.4911

Connor, G., & Fuerst, J.G.R. (2023). Linear and Partially Linear Models of Behavioral Trait Variation Using Admixture Regression. In Lynn, R. (Ed.). *Intelligence, Race and Sex: Some Controversial Issues: A Tribute to Helmuth Nyborg at 85.* London: Arktos.

Corach, D., & Caputo, M. (2022). Social injustice unveiled by genetic analysis: Argentina as a case study. *American Journal of Human Biology*, e23820.

Crutchfield, J., Keyes, L., Williams, M., & Eugene, D. R. (2022). A Scoping Review of Colorism in Schools: Academic, Social, and Emotional Experiences of Students of Color. *Social Sciences*, *11*(1), 15.

da Silva, A. K. L. S., Madrigal, L., & da Silva, H. P. (2020). Relationships among genomic ancestry, clinical manifestations, socioeconomic status, and skin color of people with sickle cell disease in the State of Pará, Amazonia, Brazil. *Antropologia Portuguesa*, (*37*), 159-176.

Dixon, A. R., & Telles, E. E. (2017). Skin color and colorism: Global research, concepts, and measurement. *Annual Review of Sociology*, *43*(1), 405-424.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., ... & Monette, G. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing*, *16*.

Francis-Tan, A. (2016). Light and shadows: an analysis of racial differences between siblings in Brazil. *Social Science Research*, *58*, 254-265.

Francis, A. M., & Tannuri-Pianto, M. (2012). The redistributive equity of affirmative action: Exploring the role of race, socioeconomic status, and gender in college admissions. *Economics of Education Review*, *31*(1), 45–55.

Fuerst, J., & Kirkegaard, E. O. (2016). Admixture in the Americas: Regional and national differences. *Mankind Quarterly*, *56*(3), 255.

Fuerst, J. G., Hu, M., & Connor, G. (2021). Genetic ancestry and general cognitive ability in a sample of American youths. *Mankind Quarterly*, *62*(1).

Fuerst, J. G. (2021). Robustness analysis of African genetic ancestry in admixture regression models of cognitive test scores. *Mankind Quarterly*, *62*(2).

Fuerst, J. G., Lynn, R., & Kirkegaard, E. O. (2019). The Effect of Biracial Status and Color on Crystallized Intelligence in the US-Born African–European American Population. *Psych*, *1*(1), 44-54.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74-78.

Hailu, S. (2018). *Skin-Tone and Academic Achievement Among 5-year-old Mexican Children*. Master's Thesis, Virginia Commonwealth University Richmond, Richmond, VA, USA.

Hall, R. E. (2020). The DuBoisian talented tenth: Reviewing and assessing mulatto colorism in the post-DuBoisian era. *Journal of African American Studies*, *24*(1), 78-95.

Hannon, L. (2014). Hispanic respondent intelligence level and skin tone: Interviewer perceptions from the American national election study. *Hispanic Journal of Behavioral Sciences*, *36*(3), 265-283.

Hannon, L., & DeFina, R. (2016). Reliability concerns in measuring respondent skin tone by interviewer observation. *Public Opinion Quarterly*, *80*(2), 534-541.

Hannon, L., & DeFina, R. (2020). The reliability of same-race and cross-race skin tone judgments. *Race and Social Problems*, *12*(2), 186-194.

Harris, A. P. (2008). From color line to color chart: Racism and colorism in the new century. The Berkeley Journal of African-American Law & Policy, *10*, 52.

Heeringa, S. G., & Berglund, P. A. (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. *BioRxiv*. doi: https://doi.org/10.1101/2020.02.10.942011

Hernández, T. K. (2015). Colorism and the law in Latin America-global perspectives on colorism conference remarks. *Washington University Global Studies Law Review, 14*, 683.

Hill, M. E. (2002). Skin color and intelligence in African Americans: A reanalysis of Lynn's data. *Population and Environment*, *24*(2), 209-214.

Hochschild, J. L., & Weaver, V. (2007). The skin color paradox and the American racial order. *Social Forces*, *86*(2), 643-670.

Hu, M., Lasker, J., Kirkegaard, E. O., & Fuerst, J. G. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych*, *1*(1), 240-261.

Huddleston, D. A., & Montgomery, L. (2010). A Scholarly Response to Shades of Black. In: Lewis, C., Jackson, J., Eds, *pp*. 63-76. *Black in America: A Scholarly Response to the CNN Documentary*.

Hunter, M. L. (2013). The consequences of colorism. In R. E. Hall (Ed.), *The melanin millennium: Skin color as 21st century international discourse* (*pp*. 247–256). Springer Netherlands.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: Springer.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Prager.

Kim, Y., & Calzada, E. J. (2019). Skin color and academic achievement in young, Latino children: Impacts across gender and ethnic group. *Cultural Diversity and Ethnic Minority Psychology*, *25*(2), 220-231.

Kim, J., Edge, M. D., Goldberg, A., & Rosenberg, N. A. (2021). Skin deep: The decoupling of genetic admixture levels from phenotypes that differed between source populations. *American Journal of Physical Anthropology*, *175*(2), 406-421.

Kirkegaard, E. O., Wang, M., & Fuerst, J. G. (2017). Biogeographic Ancestry and Socioeconomic Outcomes in the Americas: A Meta-Analysis. *Mankind Quarterly*, *57*(3).

Kizer, J. (2017). *Skin color stratification and the family: Sibling differences and the consequences for inequality* (Doctoral dissertation, UC Irvine).

Kleiber, C., Zeileis, A., & Zeileis, M. A. (2020). Package 'aer'. *R package version*, *12*(4).

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Kreisman, D., & Rangel, M. A. (2015). On the blurring of the color line: Wages and employment for Black males of different skin tones. *Review of Economics and Statistics*, *97*(1), 1-13.

Lasker, J., Pesta, B. J., Fuerst, J. G., & Kirkegaard, E. O. (2019). Global ancestry and cognitive ability. *Psych*, *1*(1), 431-459.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Social Science Genetic Association Consortium. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics*, *50*(8), 1112.

Leite, T. K., Fonseca, R. M., França, N. M. D., Parra, E. J., & Pereira, R. W. (2011). Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. *PloS one*, *6*(11), e27162.

Liu, M. M., Crowe, M., Telles, E. E., Jiménez-Velázquez, I. Z., & Dow, W. H. (2022). Color disparities in cognitive aging among Puerto Ricans on the archipelago. *SSM-Population Health*, *17*, 100998.

Marira, T. D., & Mitra, P. (2013). Colorism: Ubiquitous yet understudied. *Industrial and Organizational Psychology, 6*(1), 103-107.

Marteleto, L. J., & Dondero, M. (2016). Racial inequality in education in Brazil: A twins fixed-effects approach. *Demography*, *53*(4), 1185-1205.

Massey D., & Martin J. (2003). *The NIS Skin Color Scale*. Princeton, NJ: Princeton University Press.

Mill, R. & Stein, L. (2016). *Race, Skin Color, and Economic Outcomes in Early Twentieth-Century America*. Available at SSRN: https://ssrn.com/abstract=2741797 or http://dx.doi.org/10.2139/ssrn.2741797

Mooney, J. A., Agranat-Tamir, L., Pritchard, J. K., & Rosenberg, N. A. (2022). On the number of genealogical ancestors tracing to the source groups of an admixed population. arXiv preprint arXiv:2210.12306.

Norris, E. T., Rishishwar, L., Wang, L., Conley, A. B., Chande, A. T., Dabrowski, A. M., ... & Jordan, I. K. (2019). Assortative mating on ancestry-variant traits in admixed Latin American populations. *Frontiers in Genetics*, *10*, 359.

Rangel, M. A. (2015). Is parental love colorblind? Human capital accumulation within mixed families. *The Review of Black Political Economy*, *42*(1-2), 57-86.

Reuter, E. B. (1917). The superiority of the mulatto. *American Journal of Sociology*, *23*(1), 83-106.

Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The comprehensive R archive network*, *337*, 338.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118-137.

Roth, W. D., Solís, P., & Sue, C. A. (2022). Beyond money whitening: Racialized hierarchies and socioeconomic escalators in Mexico. *American Sociological Review*, *87*(5), 827-859.

Ryabov, I. (2013). Colorism and school-to-work and school-to-college transitions of African American adolescents. *Race and Social Problems*, *5*(1), 15-27.

Ryabov, I. (2016). Educational outcomes of Asian and Hispanic Americans: The significance of skin color. *Research in Social Stratification and Mobility*, *44*, 1-9.

Schachter, A., Flores, R. D., & Maghbouleh, N. (2021). Ancestry, color, or culture? How whites racially classify others in the US. *American Journal of Sociology*, *126*(5), 1220-1263.

Telles, E. (2004). *Race in Another America: The Significance of Skin Color in Brazil*. Princeton, N.J.: Princeton University Press.

Telles, E., Flores, R. D., & Urrea-Giraldo, F. (2015). Pigmentocracies: Educational inequality, skin color and census ethnoracial identification in eight Latin American countries. *Research in Social Stratification and Mobility*, *40*, 39-58.

Telles, E., & Paschel, T. (2014). Who is black, white, or mixed race? How skin color, status, and nation shape racial classification in Latin America. *American Journal of Sociology*, *120*(3), 864-907.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., Imai, K., and Yamamoto, M. T. (2017). *Package 'Mediation'*. https://imai.fas.harvard.edu/projects/mechanisms.html.

Thompson, M. S., & McDonald, S. (2016). Race, skin tone, and educational achievement. *Sociological Perspectives*, *59*(1), 91-111.

Thompson, W.K., Barch, D.M., Bjork, J.M., Gonzalez, R., Nagel, B.J., Nixon, S.J. & Luciana, M. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Developmental Cognitive Neuroscience 36*: 100606

Valenzuela, C. Y. (2011). Human sociogenetics. *Biological Research*, 44(4), 393-404.

Walsh, S., Chaitanya, L., Breslin, K., Muralidharan, C., Bronikowska, A., Pospiech, E., ... & Kayser, M. (2017). Global skin colour prediction from DNA. *Human Genetics*, *136*(7), 847-863.

Walsh, S., Chaitanya, L., Clarisse, L., Wirken, L., Draus-Barini, J., Kovatsi, L., ... & Kayser, M. (2014). Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Science International: Genetics*, *9*, 150-161.

Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., ... & Price, A. L. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nature Genetics*, *52*(12), 1355-1363.

Weissbrod, O., Kanai, M., Shi, H., Gazal, S., Peyrot, W. J., Khera, A. V., Okada, Y., Matsuda, K., Yamanashi, Y., Furucawa, Y., Morisaki, T., Murakami, Y., Kamatani, Y., Muto, K., Nagai, A., Obara, W., Yamaji, K., Takahashi, K., Asai, S., ... Price, A. L. (2022). Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nature Genetics*, *54*(4), 1–9.

Zou, J. Y., Park, D. S., Burchard, E. G., Torgerson, D. G., Pino-Yanes, M., Song, Y. S., ... & Zaitlen, N. (2015). Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. *Proceedings of the National Academy of Sciences*, *112*(44), 13621-13626.

CHAPTER 11

ROBUSTNESS ANALYSIS OF AFRICAN GENETIC ANCESTRY IN ADMIXTURE REGRESSION MODELS OF COGNITIVE TEST SCORES

JOHN G.R. FUERST

Abstract

Admixture regression involves the regression of a medical or behavioral trait on sub-population admixture proportions, race and ethnic group identifiers, and other explanatory variables, to measure the environmental and genetic components to trait variation across self-identified racial and ethnic groups. Several recent admixture regression studies find that proportion of African genetic ancestry has strong explanatory power for cognitive ability test scores. This paper applies stratified sub-sampling to explore the possibility that potential model mis-specification and/or variable mis-measurement underlies this empirical finding. The estimated coefficient on African ancestry is not significantly changed when the sample is stratified by low socioeconomic status/high socioeconomic status, Hispanic/non-Hispanic, Black-only/Black-White biracial, or Black/White racial group self-identification.

Keywords: admixture regression, stratified subsampling, robustness test

1. Robustness Analysis for Admixture Regression

Admixture regression involves the regression of a medical or behavioral trait on sub-population admixture proportions measured using genetic population structure analysis plus socio-economic and cultural variables. The admixture regression technique exploits the natural experiment of genetic mixing between ancestral populations to measure the magnitude of the explanatory power of sub-population-related genetic variation in explaining trait variation.

Several recent papers apply admixture regression to cognitive test scores in US data and find a negative relationship between individuals' proportions of African ancestry and their cognitive test scores, or a positive relationship between proportions of European ancestry and cognitive test scores. (Due to the adding up condition that the sum of proportional ancestries must equal one, the negative-African/positive-European linear estimates are cognate findings.) Lasker et al. (2019) use data from the Philadelphia Neurodevelopment Cohort, *N* = 7321; Kirkegaard et al. (2019) use data from the Pediatric Imaging, Neurocognition and Genetics Study, *N* =1369; Warne (2020) use a convenience sample recruited through Qualtrics , *N* = 193; Fuerst et al. (2021a) use data from the Philadelphia Neurodevelopment Cohort, *N* = 7,910; Connor and Fuerst (2021), *N* = 9972, and Fuerst et al. (2021), *N* = 10,370, both use data from the Adolescent Brain and Cognitive Development Study. All six studies find essentially the same negative relationship. Two additional papers report a negative relationship between cognitive ability and African Ancestry associated principal component scores. Guo, Lin, & Harris (2019) use data from the Add Health survey, *N* = 1,575 for the SIRE stratified Black & Hispanic Black subsample, and *N* = 7610 for the full sample. Rea-Sandin et al. (2021) use the longitudinal early steps multisite study, a low income sample, with $N = 446$ individuals with cognitive data, According to Gianna Rea-Sandin (personal communication, October 01, 2021) there was a nonsignificant negative correlation of $r = -.059$ ($p = 0.051$) between Woodcock-Johnson subtests scores and their African Ancestry associated first principal component. In this sample, there was only a small cognitive difference, at Cohen's $d = .25$, between the 225 Whites and the 118 Blacks in this sample. (Calculated from the data provided by the authors.) It is not possible to directly compare the results from these later two studies with those mentioned above, since principal component scores are used. Outside of the US, two studies also found a negative relation between African ancestry and cognitive ability as assessed by dementia screening scales. Lima-Costa et al. (2018) use data from the Bambui-Epigen (Brazil) cohort study in Brazil, $N = 1215$, while Llibre-Guerra et al. (2021) use data from a community-based sample in Cuba, $N = 585$. Owing to differences in cognitive tests and the scales used, it is not possible to compare ancestry-related effect sizes results with those from the US.

This paper implements admixture regression on stratified subsamples to explore whether the negative relationship between African proportion ancestry and cognitive test scores is due to genetic variation as posited by the admixture regression specification or potentially due to regression model mis-specification or explanatory variable mis-measurement. If the admixture regression model is correctly specified, then the coefficient estimates on the full sample should not differ from those estimated on subsamples stratified on key explanatory variables. If on the other hand there is some (unknown) source of model mis-specification or variable mis-measurement, and the admixture regression model is a false representation of the true data generating process, then stratified subsampling is likely (but not guaranteed) to produce substantially different coefficient estimates. This is the statistical logic which motivates the stratified subsampling approach used here.

2. Methods and Materials

The data, variables, and the regression model used in this analysis are detailed in Fuerst et al. (2021). Briefly, this paper uses a multi-level mixed effects three-level (site, family, individual) regression model in which family and recruitment site factors are treated as random effects. For these analyses, the dependent variable is general cognitive ability scores, derived from multi-group confirmatory factor analysis. European ancestry is set as a reference. In addition to the four non-European genetic ancestry components, the following covariates are included in the regression models: fraction_SIRE, Hispanic ethnicity (dummy coded), child immigrant status (dummy coded), family immigrant status (dummy coded), ethnic attachment, state-level racism, child reported experiences of discrimination, skin color, eye color, hair color, and general socioeconomic status (SES). See Fuerst et al. (2021) for a detailed description of all variables.

As in Fuerst et al. (2021), both the genetic ancestry and SIRE variables were left unstandardized. This was done to allow the unstandardized beta coefficients to be interpreted as the effect of a change in 100 percent ancestry/SIRE identity on one standardized unit of cognitive ability. The three eye and three hair color variables, which represent probabilities, were also not standardized. On the other hands, Ethnic attachment, state racism, discrimination factor, skin color, and SES were all standardized. See: Fuerst et al. (2021).

The samples were stratified according to the race/ethnicity of the child, given by the parents, and according to parental SES:

1) The full sample
2) High SES (greater than or equal to the sample medium score in general SES)
3) Low SES (less than the sample medium score in SES)
4) Hispanics
5) non-Hispanics
6) Black self-identification (including Black Hispanics) but excluding individuals with any other racial-identities
7) Both Black & White (including Black & White Hispanics) but excluding individuals with any other racial-identities
8) At least partially White self-identification (including Hispanic & multi-race individuals)
9) No White racial self-identification (including Hispanic & multi-race individuals)
10) At least partially Black self-identification (including Hispanic & multi-race individuals)
11) No Black racial identity (including Hispanic & multi-race individuals)

In addition to the regression results, the paper shows the partial residual plots, a form of predictor effect plot (Fox & Weisberg, 2018), with respect to African Ancestry and general cognitive ability. These show the effect of African Ancestry on cognitive ability, holding everything else in the regression model equal.

All analysis were conducted in R. For the regression models, the mixed effects regression use the lmer command from the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). The partial residual plots use the effect_plot from the jtools package (Long, 2020).

3. Research Findings

The analysis relies on a single regression specification: cognitive test scores are regressed on an intercept, four proportional ancestry variables (African, Amerindian, East Asian, and South Asian), three group identity variables (Black, Native American, and Other), one ethnic identity dummy variable (Hispanic), and eleven other explanatory variables (see Table 1).

Table 1 shows the regression results for the full sample, for the subsample consisting of individuals from families measured at below median socioeconomic status, and for the subsample consisting of individuals from families measured at or above median socioeconomic status. The regression estimates are similar in the three cases; the African ancestry coefficient (the main focus in this paper) is significantly negative in all three cases and very similar across the two subsamples. Figures 1-3 in the Appendix provide partial residual plots for these three regressions. The data distribution is much more clustered toward higher African ancestry in the low SES subsample (compare Figures 2 and 3) but there is a very negligible difference in the African ancestry coefficient estimates in the two regressions.

Recall that the questionnaire underlying the ABCD data asks separately about Hispanic ethnicity and racial identity(ies). Table 2 shows the regression estimated on two subsample choices: those who choose Hispanic ethnicity and those who do not. These are complementary subsamples: every observation is in one of the two subsamples and there is no overlap between them. The distribution of African ancestry is quite different between these two subsamples (see Figures 3 and 4) but the African ancestry regression coefficients are both significantly negative and are very similar in magnitude.

Table 1. Admixture regression on the full sample and on low socioeconomic status (SES) and high socioeconomic status subsamples. Beta coefficients are shown, with t-statistic in parentheses. SIRE = self-identified race/ethnicity; PGP = polygenic score predicted trait probability

	Full Sample *N* = 10,370	Low SES *N* = 5,170	High SES *N* = 5,200
Intercept	0.095 (1.15)	0.056 (0.49)	0.134 (1.26)
African ancestry	-0.808 (-7.15)	-0.793 (-5.57)	-0.785 (-3.76)
Amerindian ancestry	-0.878 (-7.05)	-0.970 (-6.02)	-0.826 (-3.64)
East Asian ancestry	0.170 (0.53)	-0.307 (-0.57)	0.515 (1.32)
South Asian ancestry	0.751 (1.39)	0.911 (1.35)	0.029 (0.03)
Black SIRE	0.034 (0.40)	0.025 (0.24)	0.099 (0.63)
Native Amer. SIRE	-0.007 (-0.08)	-0.023 (-0.21)	-0.016 (-0.10)
Other SIRE	-0.016 (-0.34)	-0.003 (-0.05)	-0.050 (-0.57)
Hispanic	0.044 (1.17)	0.045 (0.83)	0.049 (0.87)
US born child	0.168 (2.86)	0.162 (1.89)	0.180 (2.24)
Immigrant family	-0.106 (-4.23)	-0.179 (-4.40)	-0.051 (-1.64)
Ethnic attachment	0.033 (3.59)	0.051 (3.75)	0.013 (1.05)
State racism	0.084 (1.99)	0.104 (2.43)	0.052 (1.08)
Discrimination fact	-0.086 (-9.78)	-0.073 (-6.25)	-0.107 (-7.81)
PGP skin color	-0.011 (-0.64)	-0.031 (-1.15)	0.014 (0.62)
PGP brown eye	0.030 (0.77)	0.084 (1.27)	0.011 (0.23)
PGP intermediate eye	-0.067 (-0.45)	0.322 (1.28)	-0.283 (-1.58)
PGP black hair	-0.047 (-0.59)	0.062 (0.51)	-0.200 (-1.63)
PGP brown hair	0.030 (0.52)	0.022 (0.21)	0.072 (1.03)
SES	0.306 (25.55)	0.287 (13.18)	0.235 (7.55)

Table 2. Admixture regression on Hispanic and non-Hispanic subsamples. Beta coefficients are shown, with t statistic in parentheses. SIRE = self-identified race/ethnicity; PGP = polygenic score predicted trait probability.

	Hispanic *N* = 2,021	non-Hispanic *N* = 8,349
Intercept	-0.195 (-1.17)	0.019 (0.19)
African ancestry	-0.577 (-2.25)	-0.806 (-5.68)
Amerindian ancestry	-0.741 (-4.35)	-1.253 (-4.59)
East Asian ancestry	-1.325 (-1.47)	0.352 (1.00)
South Asian ancestry	3.756 (1.12)	0.602 (1.09)
Black SIRE	0.222 (1.50)	-0.029 (-0.27)
Native Amer. SIRE	0.045 (0.28)	-0.015 (-0.14)
Other SIRE	-0.039 (-0.69)	0.075 (0.73)
Hispanic	NA	NA
US born child	0.051 (0.57)	0.266 (3.32)
Immigrant family	0.167 (3.13)	-0.087 (-2.99)
Ethnic attachment	0.037 (1.76)	0.032 (3.05)
State racism	0.057 (1.18)	0.079 (1.71)
Discrimination fact	-0.131 (-6.89)	-0.071 (-7.19)
PGP skin color	-0.080 (-2.16)	0.012 (0.61)
PGP brown eye	0.255 (2.30)	-0.007 (-0.17)
PGP intermediate eye	0.331 (0.78)	-0.048 (-0.29)
PGP black hair	-0.012 (-0.07)	0.004 (0.04)
PGP brown hair	0.120 (0.63)	0.021 (0.35)
Socioeconomic status	0.260 (9.91)	0.317 (23.55)

Table 3 uses a subsample consisting only of individuals who choose Black as their racial identification category (and no other choices) and a separate subsample of those who choose both Black and White categories but not the Other category. Note that these are not complementary subsets but they are fully distinct; there is no overlap by construction. The Black-White biracial subsample is relatively small, *N* = 403, and this is the only case where the coefficient on African ancestry is not significantly negative. The point estimate does not differ notably from other cases, but the standard error of estimation is larger and hence the coefficient is not significant.

Table 3. Admixture regression on single-race Black and biracial Black-White subsamples. Beta coefficients are shown, with t statistic in parentheses. SIRE = self-identified race/ethnicity; PGP = polygenic score predicted trait probability

	Single-race Black *N* = 1,799	Biracial Black-White *N* = 403
Intercept	0.483 (0.69)	1.341 (1.26)
African ancestry	-0.917 (-3.41)	-0.693 (-1.45)
Amerindian ancestry	-2.459 (-2.85)	-0.784 (-0.64)
East Asian ancestry	-0.283 (-0.19)	1.325 (0.31)
South Asian ancestry	-5.759 (-1.76)	8.381 (1.28)
Black SIRE	NA	NA
Native Amer. SIRE	NA	NA
Other SIRE	NA	NA
Hispanic	0.277 (1.91)	0.202 (1.03)
US born child	0.367 (2.26)	-1.154 (-1.14)
Immigrant family	-0.282 (-3.62)	-0.093 (-0.68)
Ethnic attachment	0.079 (3.24)	0.035 (0.65)
State racism	0.041 (0.79)	0.034 (0.44)
Discrimination fact	-0.047 (-2.47)	-0.087 (-1.90)
PGP skin color	0.116 (1.70)	0.037 (0.50)
PGP brown eye	-0.600 (-1.57)	0.187 (0.80)
PGP intermediate eye	-0.625 (-0.38)	0.119 (0.10)
PGP black hair	0.137 (0.20)	0.306 (0.65)
PGP brown hair	0.177 (0.24)	-0.169 (-0.35)
Socioeconomic status	0.280 (9.39)	0.266 (4.60)

Tables 4 and 5 create schema which produces complementary subsets linked to racial identification choices. Table 4 considers the subset of individuals who choose White as one of their race identification categories, and the subset of individuals who do not choose White as one of their racial identification categories. These are complementary subsets. Note that there are no restrictions on the choices of Black or Other or Hispanic/non-Hispanic. Table 5 repeats the analogous schema for the Black category choice. See Figures 8-11 for the interesting data distributions in these four subsamples. The African ancestry regression coefficients are significantly negative in all four cases, and do not differ substantially.

Table 6 provides comparisons across the five pairs of contrasting subsamples: low SES/high SES, Hispanic/non-Hispanic, Black-only/Black-White Biracial, Individuals Including White Category in Self-identification/Not Including White Category, Individuals Including Black Category/Not Including Black Category. None of the pairs of coefficient estimates differ significantly from each other.

Table 4. Admixture regression on subsamples of individuals choosing the White category among their choices and individuals not choosing the White category among their choices. Beta coefficients are shown, with t statistic in parentheses. SIRE = self-identified race/ethnicity; PGP = polygenic score predicted trait probability

	Choosing White *N*=7,903	not choosing White *N*=2,467
Intercept	0.094 (1.03)	0.236 (0.72)
African ancestry	-0.897 (-4.19)	-0.711 (-3.72)
Amerindian ancestry	-0.692 (-4.79)	-1.210 (-4.66)
East Asian ancestry	-0.267 (-0.53)	0.488 (1.01)
South Asian ancestry	0.566 (0.59)	0.839 (1.12)
Black SIRE	0.427 (2.28)	NA
Native Amer. SIRE	0.035 (0.28)	-0.036 (-0.23)
Other SIRE	-0.276 (-0.82)	-0.022 (-0.21)
Hispanic	0.029 (0.68)	0.126 (1.33)
US born child	0.137 (1.97)	0.198 (1.73)
Immigrant family	-0.070 (-2.62)	-0.242 (-3.85)
Ethnic attachment	0.018 (1.73)	0.078 (3.77)
State racism	0.093 (2.15)	0.056 (1.09)
Discrimination fact	-0.103 (-9.73)	-0.056 (-3.40)
PGP skin color	-0.017 (-0.95)	0.005 (0.10)
PGP brown eye	0.040 (1.03)	-0.173 (-0.83)
PGP intermediate eye	-0.082 (-0.56)	-0.297 (-0.35)
PGP black hair	-0.010 (-0.10)	-0.007 (-0.02)
PGP brown hair	0.003 (0.05)	0.095 (0.27)
Socioeconomic status	0.306 (22.37)	0.271 (10.67)

Table 5. Admixture regression on subsamples of individuals choosing the Black category among their choices and individuals not choosing Black among their choices. Beta coefficients are shown, with t statistic in parentheses. SIRE = self-identified race/ethnicity; PGP = polygenic score predicted trait probability

	choosing Black *N*=2,315	not choosing Black *N*=8,055
Intercept	0.117 (0.35)	0.108 (1.25)
African ancestry	-0.831 (-3.77)	-0.858 (-5.28)
Amerindian ancestry	-1.852 (-2.87)	-0.665 (-5.15)
East Asian ancestry	-0.015 (-0.01)	0.246 (0.77)
South Asian ancestry	-3.104 (-1.08)	1.034 (1.96)
Black SIRE	-0.420 (-2.25)	NA
Native Amer. SIRE	-0.268 (-0.84)	-0.067 (-0.73)
Other SIRE	0.150 (0.23)	-0.007 (-0.14)
Hispanic	0.276 (2.54)	0.017 (0.41)
US born child	0.308 (1.98)	0.114 (1.82)
Immigrant family	-0.212 (-3.18)	-0.078 (-2.94)
Ethnic attachment	0.073 (3.36)	0.022 (2.15)
State racism	0.029 (0.55)	0.100 (2.28)
Discrimination fact	-0.056 (-3.28)	-0.103 (-9.89)
PGP skin color	0.077 (1.60)	-0.031 (-1.70)
PGP brown eye	-0.023 (-0.12)	0.062 (1.59)
PGP intermediate eye	0.886 (1.01)	-0.158 (-1.07)
PGP black hair	0.345 (1.04)	-0.109 (-1.20)
PGP brown hair	0.283 (0.82)	0.042 (0.72)
Socioeconomic status	0.283 (10.90)	0.298 (21.93)

Table 6. Comparison of African ancestry coefficient estimates across contrasting pairs of subsamples. The last column shows the *p*-value for test of equal coefficients across contrasting subsample pairs

	African ancestry coefficient	95% confidence interval		*p*-value
Full sample	-0.808	-1.030	-0.587	NA
Low SES subsample	-0.793	-1.073	-0.514	0.49
High SES subsample	-0.785	-1.194	-0.376	
Hispanic subsample	-0.577	-1.079	-0.075	0.22
Non-Hispanic subsample	-0.806	-1.084	-0.528	
Black-only subsample	-0.917	-1.444	-0.390	0.34
Black-White biracial subsample	-0.693	-1.629	0.244	
choosing White category	-0.897	-1.316	-0.477	0.26
not choosing White category	-0.711	-1.086	-0.336	
choosing Black category	-0.831	-1.263	-0.398	0.46
not choosing Black category	-0.858	-1.177	-0.539	

4. Summary

Several recent admixture regression studies find that genetic variation associated with African ancestry plays a substantial role in explaining the US Black-White test score gap: in an admixture regression model for cognitive test scores, the estimated coefficient on the proportion African ancestry is statistically significant and large in magnitude. This paper applies a set of robustness tests based on stratified subsampling to stress-test this empirical finding. If the admixture regression model used to produce this finding is an accurate representation of the true data generating process, then the estimated coefficient should not vary significantly when re-estimated on stratified subsamples. On the other hand, if the purported finding is an error due to model mis-specification or explanatory variable mis-measurement, then the estimated coefficient is likely (but not certain) to change substantially when re-estimated using selected subsamples.

This paper re-estimates the admixture regression model of cognitive test scores using five pairs of non-overlapping subsamples: low family socioeconomic status/high family socioeconomic status, Hispanic/non-Hispanic, single-Race Black/Biracial Black-White, Individuals Including White Among Their Chosen Race Categories/Not Including White Among Their Chosen Race Categories, and Individuals Including Black Among Their Chosen Race Categories/Not Including Black Among Their Chosen Race Categories. In nine of the ten subsamples, the coefficient on African ancestry is significantly negative and substantial in magnitude. In the tenth case (notable for the low number of observations) the coefficient is similar in magnitude but not statistically significant. In none of the five pair-wise comparisons of coefficient estimates on contrasting subsamples is there a substantive or significant difference between the two estimated values. This analysis provides strong support for the existing empirical finding.

References

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01.

Connor, G. and J.G.R. Fuerst (2021). Linear and partially linear models of behavioral trait variation using admixture regression. Forthcoming.

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.

Fuerst, J.G.R., Kirkegaard, E.O.W. & Piffer, D. (2021a). More Research Needed: There is a Robust Causal vs. Confounding Problem for Intelligence-associated Polygenic Scores in Context to Admixed American Populations. *Mankind Quarterly* 62: 151-185.

Fuerst, J.G.R., Hu, M. & Connor, G. (2021). Genetic ancestry and general cognitive ability in a sample of American youths. *Mankind Quarterly* 62: 186-216.

Guo, G., Lin, M. J., & Harris, K. M. (2019). Socioeconomic and genomic roots of verbal ability. bioRxiv, 544411.

Kirkegaard, E.O., Williams, R.L., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability, and socioeconomic outcomes. *Psych* 1(1): 1-25.

Lasker, J., Pesta, B.J., Fuerst, J.G. & Kirkegaard, E.O. (2019). Global ancestry and cognitive ability. *Psych* 1(1): 431-459.

Llibre-Guerra, J. J., Li, Y., Allen, I. E., Llibre-Guerra, J. C., Rodríguez Salgado, A. M., Peñalver, A. I., ... & Llibre-Rodríguez, J. J. (2021). Race, Genetic Admixture, and Cognitive Performance in the Cuban Population. *The Journals of Gerontology: Series A*.

Lima-Costa, M. F., Gouveia, M. H., Cesar, C. C., Tarazona-Santos, E., Stewart, R., Ferri, C. P., & Castro-Costa, E. (2018). Genomic African and Native American Ancestry and 15-Year Cognitive Trajectory: Bambui Study, Brazil. *Journal of the American Geriatrics Society*, 66(10), 1956-1962.

Long JA (2020). *jtools: Analysis and Presentation of Social Scientific Data*. R package version 2.1.0, https://cran.r-project.org/package=jtools.

Rea-Sandin, G., Oro, V., Strouse, E., Clifford, S., Wilson, M. N., Shaw, D. S., & Lemery-Chalfant, K. (2021). Educational attainment polygenic score predicts inhibitory control and academic skills in early and middle childhood. *Genes, Brain and Behavior*, 20(7), e12762.

Warne, R. T. (2020). Continental genetic ancestry source correlates with global cognitive ability score. Mankind Quarterly, 60(3).

Appendix: Partial Residual Plots Associated with African Ancestry for Each Admixture Regression
This appendix shows the partial residual plot and the univariate regression line associated with African ancestry for each of the admixture regression subsamples.

Figure 1: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Full Sample

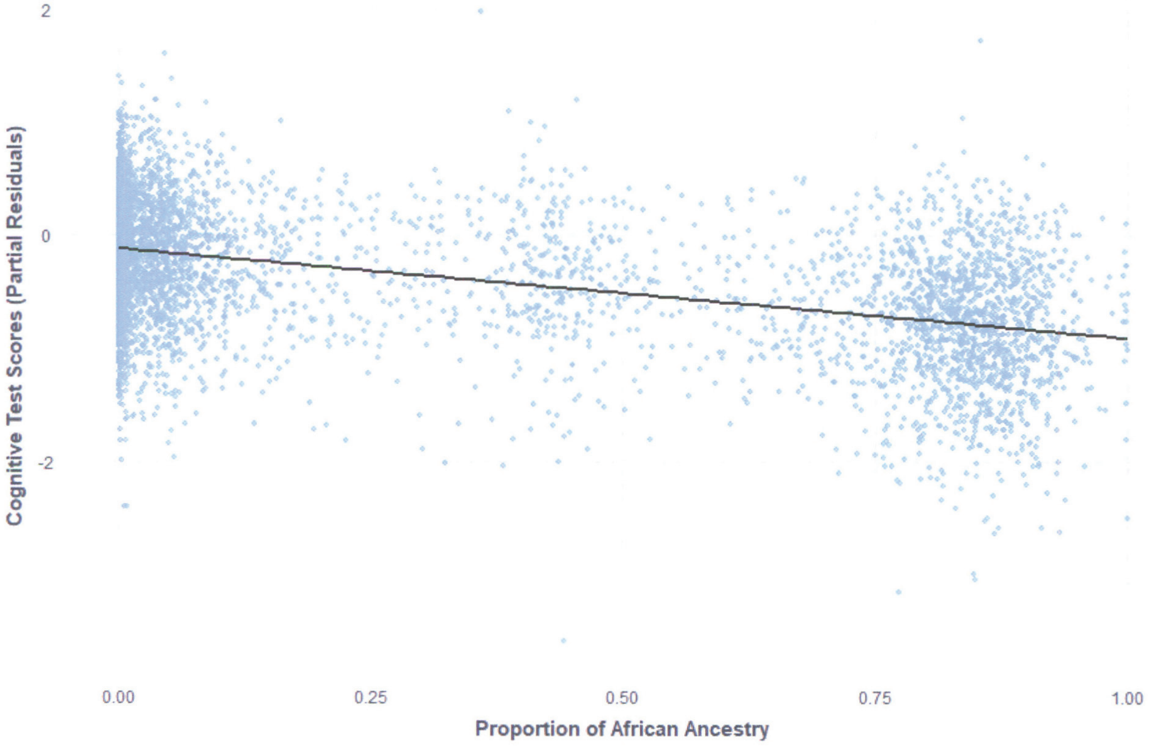


Figure 2: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Low Socioeconomic Status Subsample

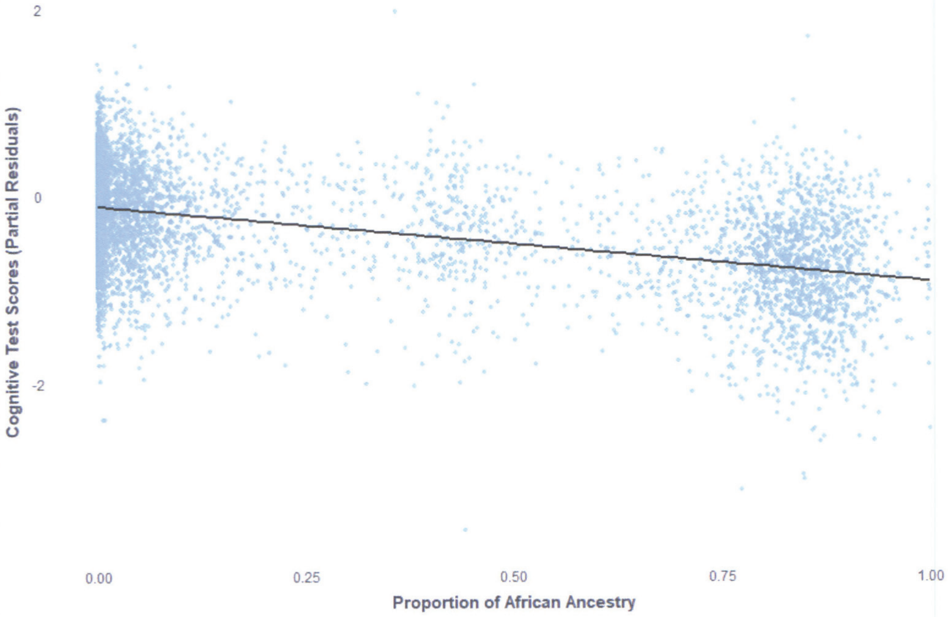


Figure 3: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: High Socioeconomic Status Subsample

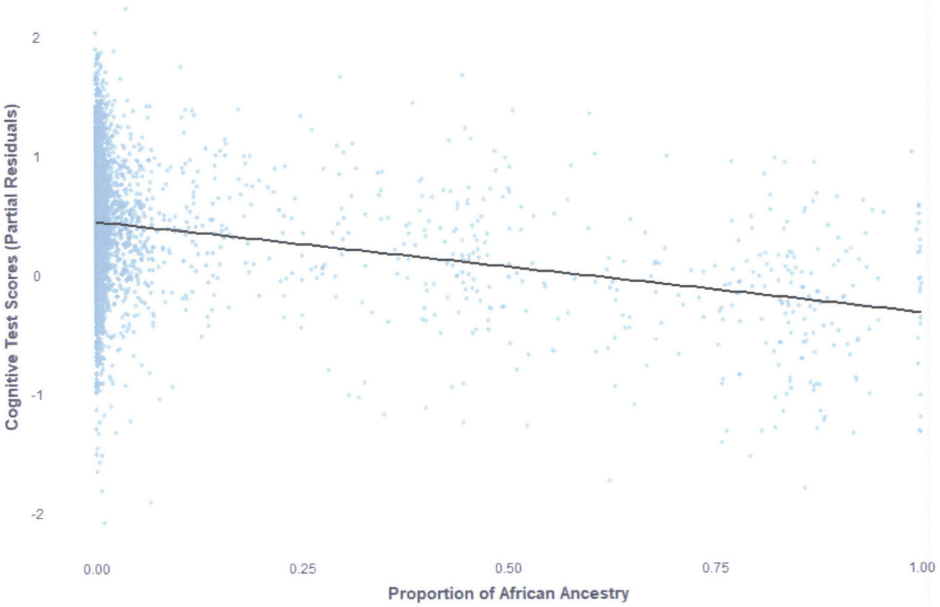


Figure 4: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Hispanic Subsample

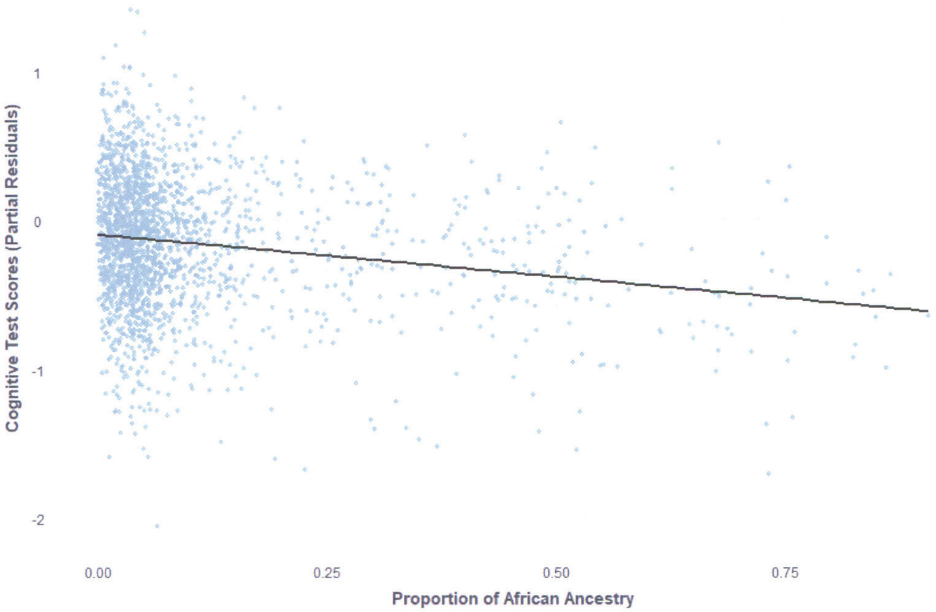


Figure 5: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Non-Hispanic Subsample

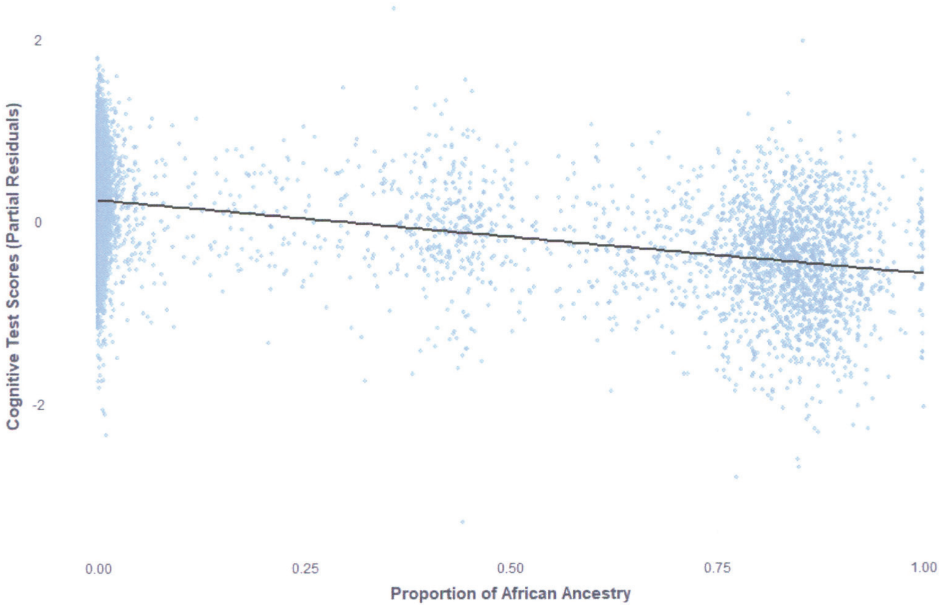


Figure 6: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Individuals Choosing Black-only Racial Identification

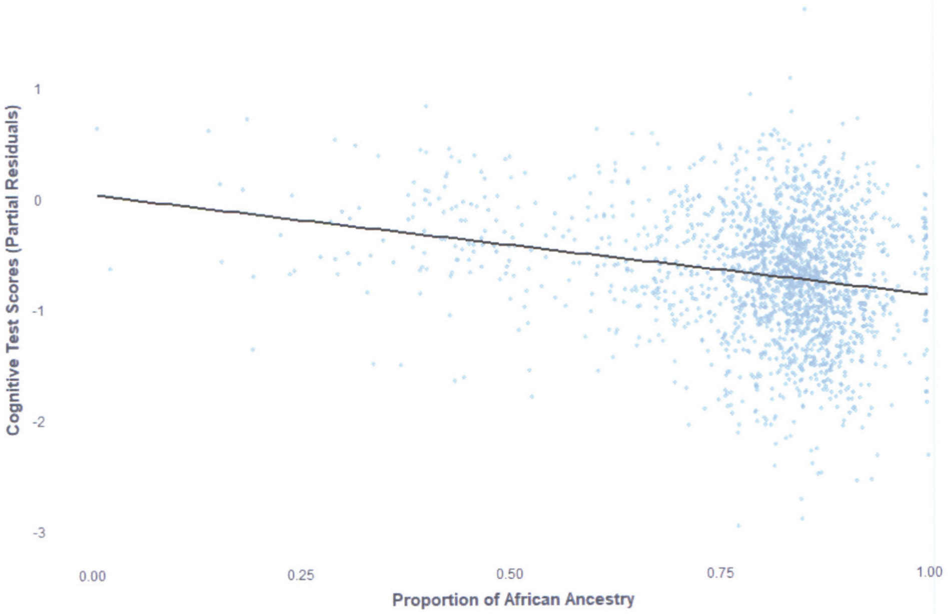


Figure 7: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Individuals Choosing Black-White Biracial Identification

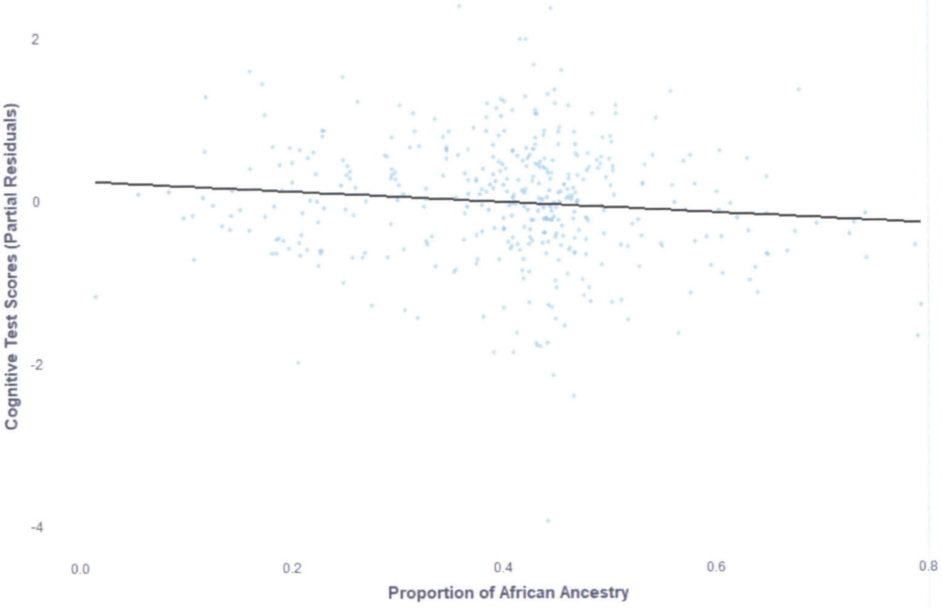


Figure 8: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Individuals Choosing White Among Their Racial Category(ies)

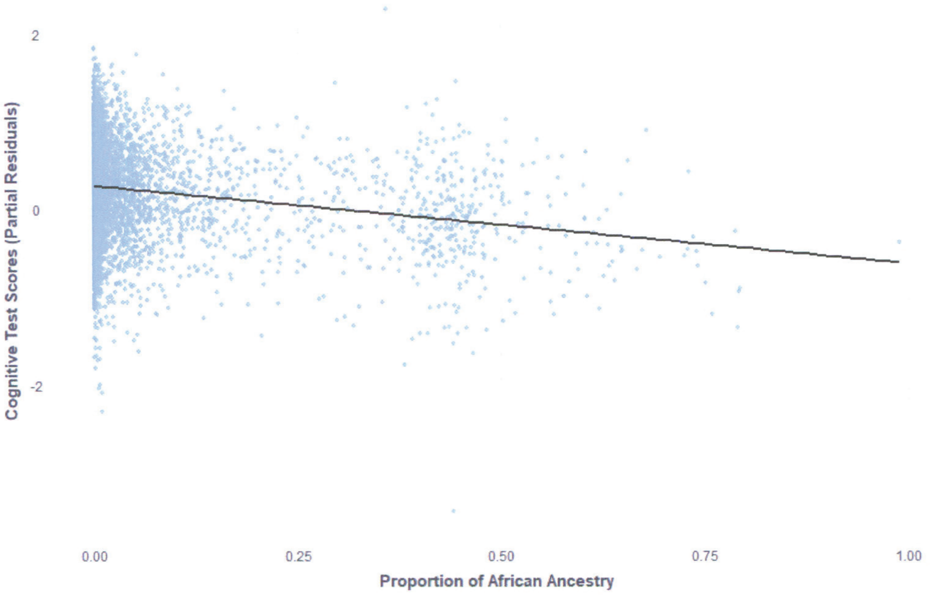


Figure 9: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Individuals Not Choosing White Among Their Racial Category(ies)

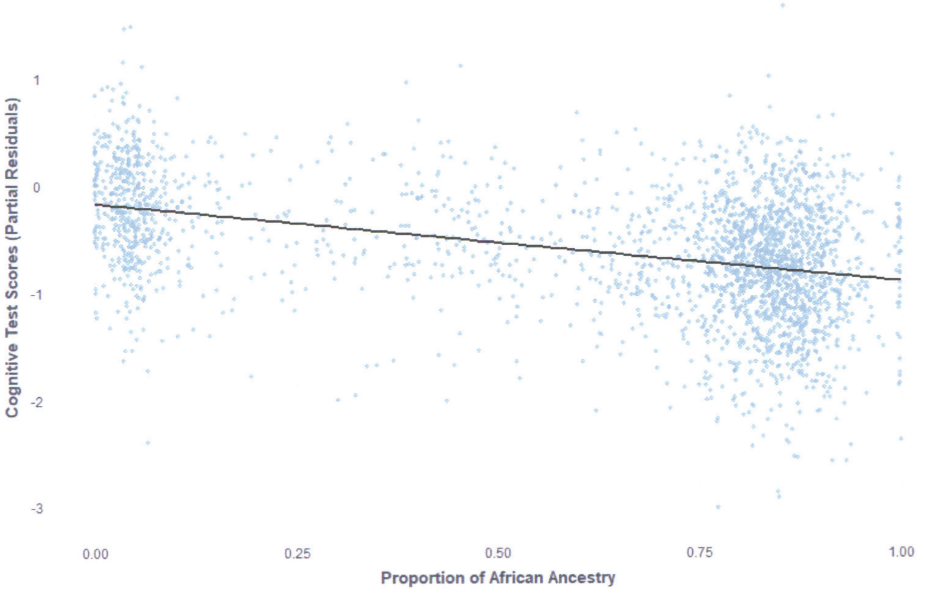


Figure 10: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Individuals Choosing Black Among Their Racial Category(ies)

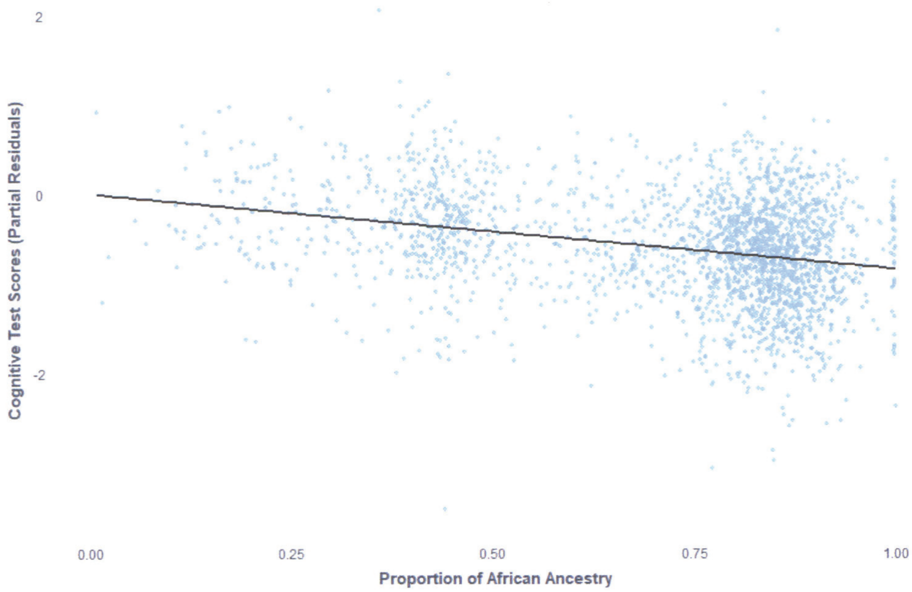
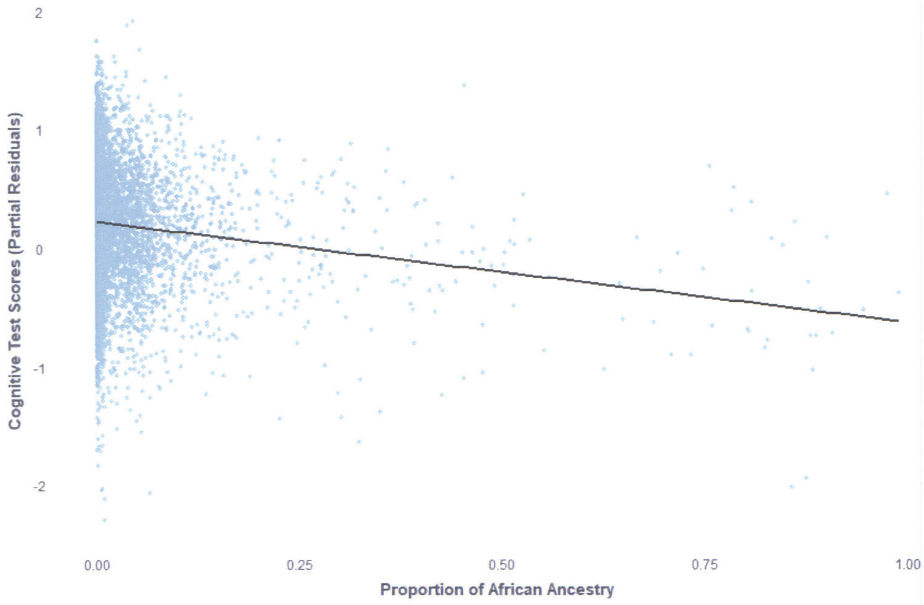


Figure 11: Partial Residuals and Estimated Regression Fit for African Ancestry in the Admixture Regression: Individuals Not Choosing Black Among Their Racial Category(ies)



PART 3:

STATISTICAL DIFFERENCES IN BRAIN STRUCTURE ACROSS BIOGEOGRAPHIC ANCESTRIES

CHAPTER 12

A GENETIC HYPOTHESIS FOR AMERICAN RACE/ETHNIC DIFFERENCES IN MEAN G: A REPLY TO WARNE (2021) WITH FIFTEEN NEW EMPIRICAL TESTS USING THE ABCD DATASET

JOHN G. R. FUERST, VLADIMIR SHIBAEV & EMIL O. W. KIRKEGAARD

Abstract

Intelligence tests are excellent predictors of school and job performance, and racial/ethnic differences in mean IQ are common. Based on five lines of evidence, Warne (2021) builds a case for partly genetic causes of differences in general intelligence (*g*) across American self/parental-identified race or ethnicity (SIRE). Based on a careful reading of Warne (2021) and the authors he cites, we generated 15 predictions flowing from a partial genetic hypothesis. These predictions concern (1) mean cognitive differences and Spearman's hypothesis, (2) measurement invariance across European genetic ancestry, (3) high within-group heritability and low shared environmentality, (4) admixture regression for *g*, (5) polygenic scores, (6) brain volume, and (7) results based on the method of correlated vectors. We used the Adolescent Brain Cognitive Development Study sample ($N = 10{,}245$) to test these hypotheses using classical and state-of-the-art statistical techniques. Decomposition of variance using twins showed that the heritability of intelligence and of brain/intracranial volume estimates were, respectively, moderate and high for both the White and the non-White subsamples, while the variance attributable to shared environment was low. Within SIRE groups, both genetic ancestry and education-related polygenic scores (eduPGS) predicted both brain volume and *g*. Moreover, brain volume was weakly but statistically significantly related to *g* ($r = .14$ to $.25$). Path and mediation analysis showed that total brain volume explained approximately 15% of the association between European ancestry and *g* and also explained approximately 8% of that between eduPGS and *g*. Finally, based on the method of correlated vectors (MCV), a positive manifold was found for genetic, brain volume, and ancestry/SIRE-related variables. We conclude that the results support the hypotheses tested and are in line with a partial genetic hypothesis.

Keywords: brain volume, *g*, ancestry, admixture, heritability, polygenic scores, SIRE

1. Introduction

IQ tests are statistically robust predictors of school and job performance, and average cognitive score differences between self/parental-identified race or ethnicity (SIRE) groups in the US such as Blacks, Whites, and Hispanics have been well documented (C. Murray, 2021; Roth et al., 2017). These differences are not due to psychometric bias, as evidenced by the finding of measurement invariance across American SIRE groups (e.g., Frisby & Beaujean, 2015; Scheiber, 2016a,b). Rather, these score differences represent real average differences in the underlying construct of general intelligence. Among intelligence researchers, there is disagreement about the degree to which differences are due to genes; notably, when surveyed anonymously, an overwhelming majority of experts, meaning people who have published on this specialist topic, attributed some part of the differences to genetics. Most of the disagreement is in the degree of the genetic contributions, varying from 0% to 100% (Rindermann et al., 2020). Determining the source of these differences is important in order to better understand and address them and their social implications (Flynn, 2018; Pesta et al., 2021).

1.1 Warne's lines of evidence

Recently, Warne (2021, based on Warne, 2020b) outlined five lines of converging evidence favoring a partially genetic model for the cause of SIRE group differences. In summary, Warne's five lines of evidence are:

1. The consistent finding of measurement invariance,
2. Finding of constraints of high within-group heritability on between-group environmentality,
3. Findings of genetics-based studies applying the admixture regression methodology,
4. Findings of genetics-based studies using education-related polygenic scores (eduPGS),
5. Large-scale, consistent support based on the method of correlated vectors (MCV).

The evidence and rationale for these five lines of evidence has been discussed by Warne (2021). Our analyses in this respect constitute an attempted replication of previous work.

1.2 Brain volume

An additional line of evidence frequently cited by proponents of a partially genetic hypothesis is the finding of average differences in brain volume (Rushton & Jensen, 2005). Warne (2021) did not find this sixth line convincing, and concluded:

> I call upon psychologists to have an open mind and to investigate the evidence for themselves, starting with the sources I have cited in this article. I also encourage social scientists to make research contributions that can address this question.

Regarding brain volume, numerous meta-analyses have confirmed that brain volume/size correlates with intelligence (Plomin & von Stumm, 2018). In a recent meta-analysis of over 194 studies, Pietschnig et al. (2022) found a correlation of $r = .24$ to .29 between brain volume and full-scale IQ. The association was found to be higher in samples which used highly *g*-loaded tests, hence tests of high cognitive complexity. As for causes, twin studies indicate that the relation between brain volume and intelligence is predominately genetic (Betjemann et al., 2010; van Leeuwen et al., 2009; Vuoksimaa et al., 2015). Moreover, recent research shows that polygenic scores for education positively relate to brain volume (Elliott et al., 2019; Judd et al., 2020), confirming that the association between brain volume and intelligence among individuals has a genetic etiology. Similarly, Lee et al. (2019) found that the association between brain volume and intelligence exists within sibships; moreover, they found that the genetic variants associated with intracranial volume overlapped with those associated with eduPGS.

A large number of studies based on autopsy, MRI, and craniometric data confirmed differences in brain volume/cranial capacity between continental ancestry groups (Beals et al., 1984; Rindermann, 2018; Rushton & Rushton, 2003), so global craniometric variation, including variation in mean brain size, is well established (Relethford, 2010). Generally, groups with their recent evolutionary past near the equator have on average smaller brains than groups which evolved further from the equator (Beals et al., 1984; Rindermann, 2018). In line with evolutionary interpretations of these geographic differences, recent studies have shown that brain morphology is related to continental genetic ancestry in admixed populations (Fan et al., 2015; Mehta et al., 2017).

Because the relation between brain volume and intelligence is predominately genetic in origin within populations, and because certain ancestry groups differ in both brain volume and intelligence, an obvious scientific conjecture is that ancestry-related differences in intelligence may be partially explained by ancestry-related differences in brain volume (Cochran & Harpending, 2009; Rushton & Jensen, 2005). Thus Rushton and Rushton (2003) claim: "[B]rain size-related variables provide the most likely biological mediators of the race differences in intelligence" (*p*. 139).

1.3 Hypotheses

In this paper, we follow Warne's call for research, and test the overarching hypothesis that American race/ethnic differences have a partial genetic basis. To do this, we first carefully read through Warne's (2021) arguments and those of the authors cited by Warne (e.g., Lasker et al., 2019; Rushton & Jensen, 2005; Warne, 2020a). From this reading, we then generated the following seven key hypotheses and tested them on a large database. When testing these hypotheses, we focused on individuals who were primarily of European, African, and Amerindian ancestry, and not those of primarily East Asian, South Asian, or Pacific Islander descent. One reason for this was because power analyses indicated that in our database the sample sizes for the latter groups were too small to detect the predicted effects for a number of the analyses. Another reason was that there were problems with reliable and interpretable East Asian, South Asian, and Pacific Islander ancestry estimates. For example, Pacific Islanders in the US are primarily of Polynesian ancestry, but there were no Polynesian samples to use as reference samples in either the 1000 Genomes or HapMap datasets. Because we did not have a Pacific Islander component, the subsample from Hawaii, which is known to contain admixed European, indigenous Pacific Islander, and East Asian subpopulations (Sun et al., 2021), was assigned East Asian ancestry in place of Pacific Islander ancestry. But because Pacific Islander and East Asian genetic ancestry are differently associated with cognitive ability (Kirkegaard et al. 2019, Table 7), including these groups might introduce confounds in our analyses; so we decided to exclude them.

1.3.1 Measurement of group differences and Spearman's hypothesis

Warne reviews the literature showing group differences on various measures. Whites generally have higher scores than Hispanics and Blacks. Our Hypothesis 1 therefore states: Between the relevant SIRE groups, there are medium to large differences in *g*, educational polygenic scores (eduPGS), and brain/intracranial volume. To determine the magnitude of differences in *g*, we employed multigroup confirmatory factor analysis (MGCFA) as recommended in the literature (e.g., Dolan, 2000). Related to this, we also tested Spearman's hypothesis, which states that group differences in cognitive ability are primarily on the *g* factor. A strong confirmation of Spearman's hypothesis has been taken as support for a genetic hypothesis because general intelligence is "related to many types of biological markers and is highly heritable" within groups (Gottfredson. 2007). Note, while between-group mean differences in *g* and brain/intracranial volume do not in and of themselves entail trait-relevant genetic differences, the existence of phenotypic difference obviously is a prerequisite for these differences to have partial genetic causes, as Warne (2021) argues.

1.3.2 Measurement invariance across European genetic ancestry

Several studies have shown that the differences across SIRE groups have the same psychometric meaning as and are caused by a subset of the same sources as those differences within SIRE groups and ancestry deciles (Lasker et al., 2019; Warne, 2021). This leads to our Hypothesis 2: Intelligence differences across European genetic ancestry will also exhibit strict measurement invariance.

1.3.3 Within-group heritability and shared environmentality

Warne (2021) correctly points out that the between-groups heritability of a trait is a mathematical function of both the quantitative genetic variation between groups and the heritability of the trait within groups (DeFries, 1972a; McClearn & DeFries, 1973). If, for example, the within-group heritability of the trait is zero, the between-groups heritability will be zero, even if there are large trait-related quantitative genetic differences between groups. So, the within-group heritability matters for Warne's (2021) hypothesis. Of course, this relation applies to quantitative traits, in which case the within-group phenotypic variance is non-zero; otherwise, the between-groups heritability becomes undefined (DeFries, 1972a, *p.* 10).

Warne also argues that when the values of within-group environmentalities (or 1 minus heritability) are modest, between-group differences are unlikely to be explained by environment-only models. This is because when environmental factors have only modest explanatory power within groups, environmental differences between groups would need to be fairly large to account for the typically observed moderate-to-large phenotypic differences. Of course, this assumes that the sources of the between-group variance are a subset of the sources of the within-group variance (Rushton & Jensen, 2005; Sesardic, 2005; Warne, 2021). With regards to environmental sources of variance, shared environmental influences, the ones that make siblings similar to one another (e.g., parental income, neighborhood and school quality, and common family experiences), are of particular interest. This is because SIRE differences in intelligence are typically hypothesized as being caused by these sorts of variables (e.g., Weiss & Saklofske, 2020) and because ancillary evidence, such as from differential regression to the mean and sibling intraclass correlation studies (Hu et al., 2019), implies that differences are due to inter-generationally transmitted factors which impact siblings similarly. So, the within-group environmentality of traits also matters for Warne's (2021) hypothesis.

It has also been hypothesized that genetic influences might be substantially lower and environmental influences substantially higher in poorer performing race/ethnic groups (e.g., Greenspan, 2022; Guo & Wang, 2002; Scarr-Salapatek, 1971; van den Oord & Rowe, 1997). If so, this would allow environmental factors that vary within groups to be more potent sources of between-group differences.

This review of the literature leads to our Hypothesis 3a: Within American SIRE groups, most of the variance in intelligence and in brain/intracranial volume is attributable to genes, while little is attributable to shared environment, and our Hypothesis 3b: American SIRE groups exhibit similar heritabilities and environmentalities.

1.3.4 Admixture regression for *g*

Warne (2021) describes admixture regression in detail and we refer the interested reader to that discussion. This method has been frequently used in genetic epidemiology over the last twenty years. It has been applied to numerous traits including obesity (Fernández & Shiver, 2004), diabetes (Cheng et al., 2012), sleep depth (Halder et al., 2015), dopamine receptor availability (Wiers et al., 2018), cigarette smoking behaviors (Choquet et al., 2021), and metabolomics (Mehanna et al., 2022). In essence, admixture regression cleverly uses the natural experiment of mating between individuals with different ancestral backgrounds to infer the environmental and genetic components to trait variation across racial and ethnic groups (Connor & Fuerst, 2022).

Because it is well recognized that there can be ancestry-correlated environmental influences (Fernández & Shiver, 2004), variables which may capture these influences are included in the regression models (Connor & Fuerst, 2022). Typically, SES is included because European ancestry is fairly consistently related to SES within admixed American populations (Kirkegaard et al., 2017). Given this, and given that research indicates that SES can statistically explain a large portion of White-Hispanic and White-Black intelligence gaps (Weiss & Saklofske, 2020), we add a general factor of SES to the regression model which is based on the education, income, finances, marriage status, and employment status of the parents, along with a rating of neighborhood crime and a sixteen-item measure of neighborhood deprivation. This compound SES variable is added in addition to SIRE, child and family migrant status, and family and recruitment-site effects, which control for, respectively, race- and ethnicity-related culture, family-migrant status, geography, and family-related factors such as the number of children in the family.

The findings from admixture regression research lead to our Hypothesis 4: The mean intelligence differences between SIRE groups are partly explained by genetic ancestry, and as a corollary, European genetic ancestry will be positively associated with *g* (relative to African and Amerindian ancestry). Note, the predictions will differ depending on the specific groups. For example, because Northeast Asians typically score somewhat above individuals of European descent, European ancestry is predicted to be negatively correlated with *g* in mixed European and Northeast Asian ancestry groups (Warne, 2020a).

1.3.5 Polygenic scores and *g*

Warne (2021) reviews the literature showing that polygenic scores (PGSs) are genetically related to *g* within SIRE groups. This leads to our Hypothesis 5: Educationally-related polygenic scores correlate positively with intelligence within SIRE groups and among siblings within families.

We subsequently used the eduPGS in two distinct ways. First, we included them in path models relating ancestry, brain volume, and *g*. When doing so, we explicitly assumed that polygenic score differences corresponded with genetic differences causative of trait differences and estimate the amount of the between-group phenotypic differences in *g* and brain volume that the eduPGS can account for. While this causal assumption may not hold for a variety of reasons (e.g., Yair & Coop, 2022), it's very plausible. This is because the eduPGS used here 1) causally explain differences within ancestries, 2) statistically explain differences between ancestries, and because 3) ancestry-associated eduPGS differences have yet to be accounted for in terms of specific forms of population structure-related confounding (Fuerst et al., 2021; Lawson et al., 2020).

Second, we used the eduPGS to estimate the effects of *g*-related genes on tests within SIRE groups (controlling for ancestry). We then correlated these effects with the magnitude of cognitive differences between SIRE groups. These later analyses avoid any assumptions regarding the meaning of eduPGS differences between ancestries.

1.3.6 Brain volume, *g*, ancestry, and eduPGS

Much has been written about the link between group differences in intelligence and brain volume (see Rushton & Jensen, 2005). This rich literature allowed us to test a series of related hypotheses that could provide indirect genetic evidence for a genetic hypothesis of differences in brain volume. Hypothesis 6a is that brain volume correlates with intelligence within SIRE groups and among siblings within families. Confirmation of this hypothesis implies that brain volume is genetically related to *g* within SIRE groups.

Hypothesis 6b is that brain/intracranial volume differences between SIRE groups are also explained by genetic ancestry. Correspondingly, Hypothesis 6c is that within SIRE groups primarily of European, African, and Amerindian ancestry, European genetic ancestry will be positively associated with brain/intracranial volume relative to African and Amerindian ancestry (see Rushton & Jensen, 2005).

Hypothesis 6d is that SES explains a portion of the relation between both ancestry and *g* and between ancestry and brain volume, and that the relations are substantially genetic in origin. This hypothesis is based on Warne's (2021, p. 488) statement that "most people live in environments that are at least partially influenced by their genes, which means that even an "environmental" variable is often partially caused by genetics." It is also based on the finding that parental SES is genetically related to child intelligence within populations (e.g., Trzaskowski et al., 2014). The finding that the relations of SES with both *g* and brain volume have a genetic basis in this sample would weaken environmental explanations which rely on SES to environmentally explain differences, as it undermines the frequently made assumption that the correlation between traits and SES is environmental in origin (i.e., "the sociologist's fallacy"; Sesardic, 2005).

In our last brain-related hypotheses, we focus on the roles that polygenic scores and brain volume play if the genetic hypothesis is correct. Hypothesis 6e is that brain volume differences are predicted by eduPGS in all SIRE groups. Hypothesis 6f is that the relation between European genetic ancestry and *g* is mediated both by eduPGS and brain volume; moreover, Hypothesis 6g is that the relation between eduPGS and *g* is mediated by brain volume. The last two hypotheses, which are tested at the same time, are logical extensions of the prediction that eduPGS and brain volume account for the association between genetic ancestry and *g*.

1.3.7 Method of correlated vectors

Warne (2021) reviews the extensive literature showing that group differences on subtests of IQ batteries are strongly related to the cognitive complexity of these same subtests. Because the cognitive complexity of IQ subtests is strongly linked to their heritability, this is indirect evidence that group differences in IQ have a genetic component. We had various variables that are hypothesized to be genetic, so our Hypothesis 7a was that there was a positive correlation between the effects on putative genetic variables across cognitive tests. Rushton (1999) showed that all genetic variables in his study correlated, so that a strong higher-order genetic factor could be computed in a factor analysis and all putative genetic variables loaded on it. This leads to our Hypothesis 7b that there would be a higher-order genetic factor resulting from the intercorrelation among genetic variables in our study.

2. Methods

2.1 Dataset

The ABCD is a collaborative longitudinal project involving 21 sites across the US. The National Institute of Health (NIH) supported ABCD is the largest longitudinal study of brain development conducted in the US to date. Approximately 11,000 9-10-year-old children were sampled at baseline, between 2016 and 2018, using a probabilistic sampling strategy. The ABCD subjects were limited to children who were fluent in English and who did not have severe medical, neurological, or psychiatric conditions. The children are broadly representative of healthy US children in this age range. Informed consent was provided by the parents.

For all analyses, we utilized the ABCD 3.01 dataset. Because our focus was on groups who are primarily of African, European, and Amerindian ancestry, we excluded any child who was identified as being either Pacific Islander or Asian.

2.2 Variables for the analyses

The following variables were used for the analyses:

1. Eleven age-corrected cognitive tests

The following cognitive measures were given at baseline: the seven NIH Toolbox® (NIHTBX) neuropsychological battery tests, NIHTBX Wechsler Intelligence Scale for Children's Matrix Reasoning, the Little Man Test (efficiency score), the Rey Auditory Verbal Learning Test (RAVLT) immediate recall, and RAVLT delayed recall.

Regarding the first seven of these, the NIHTBX neuropsychological battery was designed to measure a broad range of cognitive abilities. It consists of seven tasks which index attention (Flanker Inhibitory Control and Attention Task), episodic memory (Picture Sequence Memory Task), language abilities (Picture Vocabulary Task & Oral Reading Recognition Task), executive function (Dimensional Change Card Sort Task & Flanker Inhibitory Control and Attention Task), processing speed (Pattern Comparison Processing Speed Task), and working memory (List Sorting Working Memory Task) (Akshoomoff et al., 2014; Weintraub et al., 2014; Thompson et al., 2019).

For the seven NIHTBX subtests, the ABCD precomputed age-corrected scores were used. For the remaining four tests, scores were adjusted for age by applying regression with a cubic spline of age to the full sample. Subsequently, scores for all eleven tests were standardized.

2. General intelligence (*g*) scores

Multi-group confirmatory factor analysis (MGCFA) was previously conducted based on the 11 cognitive tests noted above (Fuerst, Hu, & Connor, 2021). We repeated the analysis using the same model and specifications. When doing so, we first assessed if outliers and missing data had any impact, and also if our results remained robust after imputation of missing data, removal of outliers, and adjustments for age and sex. Next, we conducted both exploratory factor analysis and multi-group confirmatory factor analysis. A three-broad-factor model (memory, complex cognition, and executive function) with a general factor (*g*) fit the data well. As previously found, strict measurement invariance held across the SIRE groups (and across sex groups and age measured in months). The complete results are detailed in Appendix A.

In the best fitting and additionally most parsimonious model for SIRE group differences, *g* alone explains the differences. *g* factor scores from this MGCFA model were standardized ($M = 0.00$; $SD = 1.00$) in the full sample of 10,245 children.

3. NIH Toolbox® (NIHTBX) Cognition Battery

As the *g* scores used for the admixture regression analyses were dependent on MGCFA model specification, we additionally used the age-corrected NIH Toolbox® Cognition Battery (NIHTBX) summary composite scores ("nihtbx_totalcomp_agecorrected") precomputed by NIH as an alternative measure of general cognitive ability. Our choice was made for the sake of replicability. The NIHTBX was normed for samples between ages 3 and 85; tasks correlate highly with comparable ability assessments (Weintraub et al., 2014). The NIHTBX cognition battery had previously been found to be measurement invariant across American Black, Hispanic, and White SIRE groups (Lasker et al., 2019). Prior to analysis, we standardized scores on the full sample.

4. Unadjusted and adjusted brain and intracranial volume

ABCD provides summary brain ("smri_vol_scs_wholeb") and intracranial volume ("smri_vol_scs_intracranialv") variables which show volume in mm^3. We standardized these unadjusted variables.

We additionally created sex-, age-, and MRI assessment site-adjusted brain variables to use in the path analysis models and biometric analyses. We adjust for collection site to account for differences in MRI protocol across collection sites. Brain and intracranial volume are adjusted for sex and for the cubic spline of age-, height-, and site-related fixed effects. These adjusted brain variables were then standardized.

5. Child US born & immigrant family

Parents were asked if the child was born in the "United States". This variable is recoded as "1" for "United States" and "0" for all other responses. Parents were also asked if anyone in the child's family, including maternal or paternal grandparents, was born outside of the United States. This variable is also coded as "1" for "Yes" and "0" for all other responses.

6. General socioeconomic status (SES)

A general factor of SES was computed by applying principal components analysis (PCA) to seven indicators of SES. To analyze the data, we used the PCAmixdata R package (Chavent et al., 2014), which allows handling mixed categorical and continuous data. The first unrotated component explained 42% of the variance, indicating a strong general factor. The loadings on the first PCA factor for the seven indicators were: financial adversity (.31), area deprivation index (.49), neighborhood safety protocol (.31), parental education (.54), parental income (.66), parental marital status (.43), and parental employment status (.23). We note that this variable was standardized in the full sample. Details are provided in Appendix B.

7. SIRE fraction & Hispanic

Based on the 18 questions inquiring about the child's race, we created four dummy SIRE variables: Black, White, Native American, and Not Otherwise Classified (NOC). The latter category includes those identified as: "Other race," "Refused to answer," or "Don't know". These variables were then converted into interval variables, calculated as the value selected for each of the four groups (0 or 1) over the total number of responses (1 to 4). As a result, individuals were assigned four SIRE fractions (frac_White SIRE, frac_Black SIRE, frac_Native_American SIRE, & frac_NOC SIRE), each ranging from 0 to 1. For example, a Hispanic individual who chose both White and Black SIRE receives a ½ weighting on each of the associated group variables; an individual who chose all three of Black, Native American and White SIRE has 1/3 weight for each. This ensures that the sum across the group variables equals one for every individual. The White category is used as the base or benchmark group and the associated variable is dropped from the regression. This SIRE coding is used because interval SIRE variables had previously been found to be most predictive when included in models alongside genetic ancestry (Kirkegaard et al., 2019). These variables were left unstandardized, allowing the unstandardized beta coefficients for SIRE fraction to be interpreted as the effect of a change in 100 percent SIRE identity on one standardized unit of the dependent variable.

We also created a dummy variable for Hispanic ethnicity, coded as "1" for "Hispanic" and "0" for non-Hispanic.

8. Height

ABCD provided a summary variable ("anthroheightcalc"), which records the child's height (in inches) based on an average of up to three measures. We divide the values by 12 to give height in feet and then standardize the results in the full sample.

9. Genetic ancestry

Subjects were genotyped using Illumina XX, with 516,598 variants directly genotyped and surviving the quality control carried out by the data provider. We used the 3.0 release of the genotypic dataset, which also includes an edition with imputed variants using TOPMED and the Eagle 2.4 software. All our work was done on build 38. Files in hg17/37 were lifted to hg38 using liftOver (https://github.com/sritchie73/liftOverPlink) and the GRC chain file at ftp://ftp.ensembl.org/pub/assembly_mapping/homo_sapiens/ (GRCh37_to_GRCh38.chain.gz).

Before global admixture estimation, we applied quality control analyses using plink 1.9. We used only directly genotyped, bi-allelic, autosomal SNP variants ($N = 494{,}433$ and 493,196, before and after lifting, respectively). We pruned variants for linkage disequilibrium at the 0.1 R^2 level using plink 1.9 (--indep-pairwise 10000 100 0.1), as recommended in the admixture documentation (https://vcru.wisc.edu/simonlab/bioinformatics/programs/admixture/admixture-manual.pdf). This variant filtering was carried out in the reference population dataset to reduce bias due to sample representativeness. After pruning, we were left with 99,642 variants. To ensure a reasonable balance of populations in the estimation dataset, we merged the target samples from ABCD with reference population data for the populations of interest. We desired a k=5 solution (European, Amerindian, African, East Asian, and South Asian), so we merged with relevant samples from the 1000 Genomes database and from the HGDP database. We used all populations in the 1000 Genomes database and from the HGDP database, except Adygei, Balochi, Bedouin, Bougainville, Brahui, Burusho, Druze, Hazara, Makrani, Mozabite, Palestinian, Papuan, San, Sindhi, Uygur, and Yakut. These reference populations were excluded because they were overly admixed or because, in the case of Melanesians and San, the individuals in the ABCD sample lacked significant portions of these ancestries.

Because the estimation sample would still be highly skewed towards European ancestry when using this joint sample, we used repeated subsetting to achieve balance. Specifically, we split the ABCD target samples into 50 random subsets, each with approximately 222 persons, and merged them one at a time with the reference data, followed by running admixture k=5 on each merged subset. We verified that these subsets produced stable results by examining the stability of the estimates for the reference samples. There was very little variation across runs, e.g., for the reference sample with the most variance (European, NA12342), the mean estimate was 98.3% with *SD* = 0.17% across the 50 runs. Because the Admixture software does not label the resulting clusters, we used five reference samples to index the populations so the data would be merged correctly. In no case did this produce inconsistencies.

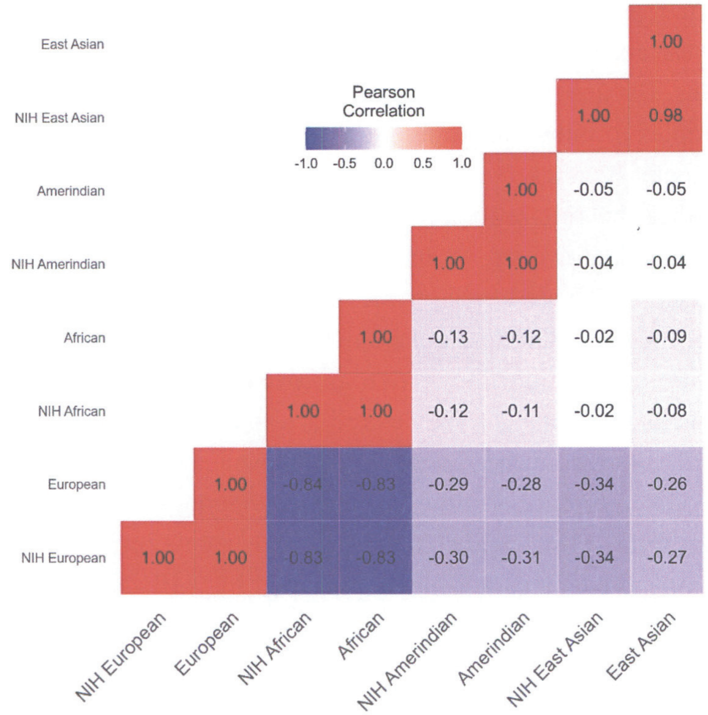
A reviewer stated that the reliability of our admixture estimates was so low that it would invalidate all our conclusions. However, we had validated our genetic ancestry percentages against those in the ABCD dataset, which are based on a $K = 4$ solution (African, European, East Asian, and American ancestry). For these, ABCD researchers used 1000 Genomes populations as the reference samples and fastStructure as the algorithm (Hatton, 2018). Note, we computed our own genetic ancestry percentages for two reasons: 1) We desired a $K = 5$ solution to separate South Asian and European ancestry; 2) We were concerned with ABCD's Amerindian admixture estimates as these were computed using only reference samples from 1000 Genomes (which had, for Amerindian ancestry, only Mexicans from Los Angeles and Peruvians from Lima, Peru). Figure 1 shows the correlation matrix for our percentages and those provided by ABCD, and it can be seen that the Amerindian, African, and European ancestry percentages are perfectly correlated and the East Asian percentages are nearly perfectly correlated ($r = .98$). So, both sets of ancestry estimates can be considered virtually perfectly reliable.

By design, the estimated ancestry proportions for an individual i, $A_{i1}, A_{i2}, \dots, A_{i5}$, are non-negative and sum to one for every individual, using the following formula:

$$\sum_{h=1}^{5} A_{ih} = 1 \; for \; every \; i$$

In order to use all ancestry variables in a regression context, one category must be chosen as the base group and the associated variable dropped from the regression, otherwise the unit-sum condition across the variables produces a singularity in the regression model. When using multiple ancestries, we used the European ancestry category as the base, because it has the largest number of nonzero observations. We also used the European ancestry variable as the single ancestry for ancestry by outcome predictor effect plots. This was because these effect plots are with respect to a single ancestry predictor.

These ancestry variables were left unstandardized. This allows the unstandardized beta coefficients for genetic ancestries to be interpreted as the effect of a change in 100 percent ancestry on one standardized unit of the dependent variable.



	NIH European	European	NIH African	African	NIH Amerindian	Amerindian	NIH East Asian	East Asian
East Asian								1.00
NIH East Asian							1.00	0.98
Amerindian						1.00	-0.05	-0.05
NIH Amerindian					1.00	1.00	-0.04	-0.04
African				1.00	-0.13	-0.12	-0.02	-0.09
NIH African			1.00	1.00	-0.12	-0.11	-0.02	-0.08
European		1.00	-0.84	-0.83	-0.29	-0.28	-0.34	-0.26
NIH European	1.00	1.00	-0.83	-0.83	-0.30	-0.31	-0.34	-0.27

Figure 1. Matrix of correlations between genetic ancestry percentages used in this paper (e.g., European) and those provided in the ABCD dataset (e.g., NIH European)

10. MTAG-eduPGS & poly-eduPGS

For the main analyses, we created two eduPGS variables. First, we used the genome-wide association study (GWAS) results from Lee et al.'s (2018) meta-analysis, based on European-descent individuals. Specifically, the multi-trait analysis of genome-wide association study (MTAG) eduPGS SNPs ($N = 8{,}898$ variants in the ABCD sample). MTAG is a method for analyzing statistics from genome-wide association studies (GWAS) on different but genetically correlated traits (e.g., education and intelligence). Lee et al. (2018) applied MTAG to four traits: education (N = 1,131,881), intelligence ($N = 257{,}841$), hardest math class taken ($N = 430{,}445$), and mathematical ability ($N = 564{,}698$); we used Lee et al.'s MTAG results for education. This predictor, denoted MTAG-eduPGS, was used because it was previously shown to predict *g* in European, Hispanic, and African-American populations (Lasker et al., 2019) and because it was assessed for some commonly proposed forms of population-structure-related confounding (e.g., biasing effects of discovery population-specific variants and of derived versus ancestral variant status, biasing effect of SNP sign discordance between populations, and the discrepancy between population-GWAS and within family coefficients) (Fuerst et al., 2021).

Second, we created eduPGS, denoted poly-eduPGS, using the new PolyFun predictor detailed by Weissbrod et al. (2022) (using $N \sim 750{,}000$ variants in the ABCD sample). Weissbrod et al. (2022) trained this predictor on 334,353 unrelated individuals of British descent from the UK Biobank. The relevant trait was having a college education or not. This predictor uses genome-wide functionally-informed fine-mapping to precisely estimate causal effects. In doing so, this method is known to circumvent linkage disequilibrium differences, which can confound trans-ancestral comparisons (Weissbrod et al., 2020, 2022). Moreover, as the training sample only includes the UK Biobank data, it is relatively homogeneous; this characteristic should reduce population-structure-related confounding. While this predictor does not provide an accurate estimate of the effect size, it provides a precise estimate of causal effects, at least given a sufficiently large training sample. Weissbrod et al. (2022) provide the SNP set and SNP weights, based on the UK Biobank, and the code for generating this eduPGS. Weissbrod et al. (2022) also created additional predictors called PolyPred and PolyPred+. These combined Polyfun with other predictors that do not use functionally-informed fine-mapping (e.g., BOLT-LMM). The combined predictors, which can incorporate non-European training samples, have been shown to increase the predictive validity of the polygenic scores. We did not use the combined predictors because we wanted a predictor based on fine-mapping only, because we were particularly concerned with confounding due to differences in linkage disequilibrium.

While an updated meta-analysis of educational GWAS has since been published (Okbay et al., 2022), only about 10k variants have been made public. This is because results based on the 23andME sample, which accounted for most of the sample size increase since Lee et al.'s (2018) meta-analysis, were not published. Moreover, this SNP set is made available by the Social Science Genetic Association Consortium (SSGAC), and the SSGAC currently proscribes the use of eduPGS for making comparisons between ancestry groups. So, it is not possible to assess population-related confounding with these eduPGS. Given this, the eduPGS above are the best currently available ones for the purposes of the present analyses.

11. First 20 genetic principal components

For several of the analyses involving eduPGS, we took population-structure-related effects into account by controlling for the first twenty ancestry principal components generated by PLINK v1.90b6.8. We used all imputed SNPs to compute these PCs.

12. Pseudo eduPGS

Using PLINK v1.90b6.8, we selected ten random sets of 8,898 variants. We randomly assigned the MTAG-eduPGS's beta weights to these sets of SNPs. We used these individually and we also averaged these to create averaged pseudo eduPGS weights.

2.3 Analyses

2.3.1 Analyses related to measurement of group differences and Spearman's hypothesis

We computed means, standard deviations, and effect sizes for *g*, the two eduPGS, and brain/intracranial volume. For effects sizes, we computed Cohen's *d* values. This commonly used metric is computed as the difference between the group means over the pooled standard deviation for the two groups.

2.3.2 Analyses related to measurement invariance

We repeated the multi-group confirmatory factor analysis (MGCFA) of Fuerst et al. (2021) as detailed in Appendix A. We additionally ran local structural equation models (LSEM) to determine if measurement invariance also held with respect to European genetic ancestry deciles in addition to the four SIRE groups. This method was applied to the 10,245 White, Hispanic, Black, and Other individuals with European genetic ancestry, *g*, and brain volume scores.

The LSEM method has been detailed by Hildebrandt et al. (2016) and was previously applied to genetic ancestry and intelligence by Lasker et al. (2019) and Lasker (2019) using two different samples. As with Fuerst et al. (2021), when conducting confirmatory factor analysis we elected to use a higher-order factor (HOF) model based on theoretical grounds (Decker, 2021; Hood, 2010) and on the absence of empirical disconfirmation of the HOF model (Murray & Johnson, 2013).

To assess invariance across European genetic ancestry, the following model-fit indices are used: chi-square, comparative fit Index (CFI), root mean square error of approximation (RMSEA), and Bayesian information criterion (BIC). CFI estimates the discrepancy between the proposed model and the null model: larger values indicate better fit. RMSEA estimates the discrepancy related to the approximation, or the amount of unexplained variance (residual), or the lack of fit compared to the saturated model: smaller values indicate better fit. BIC is a comparative measure of fit used in the comparison of two or more models; it evaluates the difference between observed and expected covariances: smaller values indicate better fit.

We selected European ancestry as the ancestry variable because there was no sparsity of data along this ancestral component.

2.3.3 Analyses related to within-group heritability and shared environmentality

A same-sex monozygotic and dizygotic twin sample was nested within the ABCD sample. Details for this twin sample are provided by Iacono et al. (2018). We used the *SEM* feature from the Lavaan (Rosseel et al., 2012) and the umx (Bates et al., 2019) packages to decompose the variance using an A (additive genetic), C (shared environment), and E (non-shared environment) model. For both NIHTBX/*g* and brain/intracranial volume, this ACE model fits better than an alternative ADE model, which replaces the C component with D (dominance) or an alternative AE model, which excludes shared environment. This assessment is based on a combination of the following fit indices: root mean square error of approximation (RMSEA), Tucker–Lewis index, and χ^2 / df. For the SEM, we used the theoretical genetic correlation of 1.0 for MZ and 0.5 for DZ. The correlations for C and E, for both kinship classes, were set to 1 and 0, respectively. In describing variance components, we adopted Chen et al.'s (2022) guidelines, which indicate that low, moderate, and high genetic variance components correspond to less than 30% of the variance, between 30% and 60%, and greater than 60%, respectively.

For these analyses, twin pairs were identified and classified using the ABCD's precomputed family relationship and genetically inferred zygosity variables. Variance component estimates were computed for the eleven cognitive subtests, *g* scores, the NIHTBX composite scores, unadjusted brain volume, unadjusted intracranial volume, adjusted brain volume, and adjusted intracranial volume. For these analyses we did not use imputed data, except in the case of the RAVLT memory trials. Specifically, in this case, we imputed missing data for, and based on only, the six RAVLT trials (pea_ravlt_sd_trial_i_tc, pea_ravlt_sd_trial_ii_tc, pea_ravlt_sd_trial_iii_tc, pea_ravlt_sd_trial_iv_tc, pea_ravlt_sd_trial_v_tc, and pea_ravlt_ld_trial_vii_tc). This was done because the short-term memory scores represent the average of the first five trials and so missing data need to be handled.

To allow heritabilities for intelligence to be compared with previously published results (e.g., Pesta et al., 2020, 2023), for these biometric analyses we defined "Black" as including any non-Hispanic individual identified as being African American. Thus "Black" was defined broadly and includes multi-racial individuals. The "Other" group, correspondingly, was defined as everyone who was not White, Hispanic, or Black as just defined. Because our primary comparison, for the biometric analyses, was between the White and the combined Hispanic, Black, and Other subgroups, this decision did not affect the interpretation of our results. This definition of "Black" differs from that used for the admixture regression analyses discussed below. In those, the "Black" subgroups included only those individuals who were marked as being Black. The latter was done for the purposes of minimizing possible sociocultural variation which may confound admixture analyses.

We report the a', c', and e' estimates, which are the square roots of the A, C, and E estimates, because these represent the correlation between traits and, respectively, genes, shared environment, and non-shared environment. These values are directly relevant to Warne's (2021) argument, because the magnitude of environmental variation within populations needed to cause phenotypic differences between populations is calculated as the phenotypic difference divided by the environmental-trait correlation.

2.3.4 Analyses related to admixture regression for *g*

The rationale for admixture regression analysis has been described in detail previously (Connor & Fuerst, 2022; Kirkegaard et al., 2019; Lasker et al., 2019). For these analyses, we used the pooled data with the ABCD baseline sample and the twin sample. We first subset to cases which had admixture estimates, *g* scores, and brain volume, leaving 10,245 cases. We then imputed missing data, yielding the same number of cases. We ran a series of full-sample and also SIRE-stratified (White, Black, Hispanic, and Other) regression analyses so as to control for potential SIRE-related environmental confounds. This was because "stratifying on SIRE has the potential benefits of reducing heterogeneity of these non-genetic variables and decoupling the correlation between genetic and non-genetic factors," since SIRE "acts as a surrogate to an array of social, cultural, behavioral, and environmental variables" (Fang et al., 2019, *p*. 764). For these analyses, the SIRE subgroups were defined using the ABCD race_ethnicity variable. There are four mutually exclusive groups in terms of SIRE categorization – non-Hispanic White (*White*) only, Hispanic of any race (*Hispanic*), non-Hispanic Black (*Black*) only, and any other (*Other*).

These were regression analyses with *g* as the dependent variable and ancestries as the key predictors. These models additionally included: child US born, family immigrant status, sex, age, and SES variables. For the total-sample analyses, the Hispanic and SIRE fraction variables were also included. Likewise, the SIRE fraction variables were included for the Hispanic subsample analyses.

As advised by Heeringa and Berglund (2020), we used a linear mixed-effects model rather than ordinary least squares. This involved partially decomposing the residual term into linear random effects components linked to the data collection site identifiers and same-family identifiers within the sample. This allows for the possibility of error term correlations within data collection sites or within families with multiple tested individuals (see Heeringa & Berglund, 2020). As Heeringa and Berglund (2020) note, this specification replicates that which is used by the ABCD Data Exploration and Analysis Portal (DEAP). Thus, the use of this multilevel model also aids in replication. For the mixed-effects regression models, we employ the *lmer* command from the lme4 package (Bates et al., 2015).

In addition to the regression results, we depicted the partial residual plots for European ancestry with *g* as the dependent variable. Partial residual plots are a form of predictor effect plot (see Fox & Weisberg, 2018). These plots show the effect of European ancestry on *g* while holding everything else in the regression model constant. These plots were created using the *effect plot* command from the jtools package in R (Long, 2020). This command uses the output from the mixed-effects regression models for genetic ancestry *g*.

2.3.5 Analyses related to polygenic scores and *g*

These were regression analyses with *g* as the dependent variable and eduPGS as the main predictors. These models additionally included: sex, the cubic spline of age, and the first twenty genetic PCs. For these analyses we used a linear mixed-effects model as described above. We ran the models both on the full sample and the SIRE-stratified subsamples. Associations between intelligence and eduPGS may be confounded by demographic factors (Zaidi & Mathieson, 2020). As such, we also run sibship analyses to assess the effects of MTAG-eduPGS and Poly-eduPGS on intelligence within families. We then compare the magnitude of these within-sibship effects to the effects found in the full sample. We use the model detailed by Howe et al. (2021, 2022) and a modified version of the R-code provided by these authors (https://github.com/LaurenceHowe/SiblingGWAS). This model gives the effect of the predictor on the criterion within sibships (denoted PXCi) and between sibships (denoted PFXi) in addition to within the full sample. The sibship analyses are based on families with multiple children, with MZ twins excluded. We use two criteria, *g* and NIHTBX composite scores. Following Howe et al. (2021, 2022) all analyses include age and sex as covariates. We also include the first 20 PCs as covariates, except when using European ancestry as the predictor. When using European ancestry as the predictor, we instead include East and South Asian ancestry as covariates. To maximize power for these analyses we use all available White, Hispanic, Black, or Other individuals in a pairwise deletion fashion. To be clear, we do not run within-sibship GWAS analyses, rather we determine if population-wide eduPGS predict *g* among sibships.

2.3.6 Analyses related to brain volume, *g*, ancestry, and eduPGS

2.3.6.1 Brain volume and g

These are regression analyses with *g* as the dependent variable and brain volume, adjusted for sex, the cubic spline of age, and height as predictors. A second set of regression models added the first twenty genetic PCs. For these analyses we used a linear mixed-effects model as described above. We ran the models both on the full sample and the SIRE-stratified subsamples. We additionally ran the within-sibship analyses to assess the effects of brain volume on intelligence within families.

2.3.6.2 Brain/ Intracranial volume, SIRE, and genetic ancestry

These were admixture regression analyses with unadjusted brain or intracranial volume variables as the dependent variable and genetic ancestries as the main predictors. The baseline model included SIRE variables but not genetic ancestries. These models, which were run on the full sample of 10,245 children, additionally included: child US born, family immigrant status, sex, age, height, and SES variables.

2.3.6.3 Brain/intracranial volume within SIRE groups

These analyses repeated the one discussed immediately above but on the SIRE subsample. As such, SIRE fraction variables were included only for the Hispanic subsample analysis.

2.3.6.4 SES, brain volume, & g among adopted and biological children

We investigated the relation between SES, brain volume, and *g* by taking advantage of the small adoption sample in the ABCD dataset. We defined adopted children as children for whom both parents were reported to be adoptive parents. There were 126 such children and there were 6400 biological children. We compare the associations for adoptive children with those for biological children. We defined biological children as children for whom both parents were reported to be biological parents. We fit a regression model with *g* and, alternatively, brain volume as the dependent variables. In these models, we included European ancestry, SIRE, the cubic spline of age, sex, child US born, family migrant status, and site- and family-fixed effects. European ancestry was included in this model because, as shown in Figure S14a of the Supplementary File 1, it was also linearly related to *g* in the sample of adopted children.

2.3.6.5 Brain volume and eduPGS

These were linear mixed-effects regression analyses with unadjusted brain volume as the dependent variable and eduPGS as the main predictors. These models additionally included: sex, age, and the first twenty genetic PCs.

2.3.6.6 Mediation by brain volume

While path models cannot prove causal assumptions, they can provide estimates of effects given these causal assumptions, for example, that increased brain volume causes increased *g,* instead of vice versa, or that eduPGS differences cause *g* differences, instead of being spuriously related (Bollen & Pearl, 2013). As such, we model the relationship between genetic ancestry, polygenic scores, adjusted brain volume, and *g* using the *lavaanPlot* command from the lavaan package in R (Long, 2020). Because we used cross-sectional data, the causal assumption imbedded in the brain volume x *g* path cannot be verified. Yet this path, with mental ability being dependent on brain matter, is theoretically well grounded. In these models, brain volume is adjusted for the effects of sex, the cubic spline of age, and height.

We additionally ran mediation analysis using the mediation R package (Tingley et al., 2013). This package estimates the proportion of the effect that is directly causally mediated using a different set of assumptions than used in path analysis (see Imai et al., 2014). For these models we used the poly-eduPGS, because it is supposed to estimate causal effects and be less confounded by population structure. In these models both the outcome and mediator variables were residualized for demographic confounds.

2.3.7 Analyses related to the method of correlated vectors

2.3.7.1 Method of correlated vectors

Finally, we applied Jensen's method of correlated vectors (MCV) to examine the patterns of the genetic, brain volume, and admixture variables' effects on cognitive test scores. To be clear, we do not use MCV to test Spearman's hypothesis, the hypothesis that group differences are mostly on *g*. For that, we use multi-group confirmatory factor analysis (MGCFA) as recommended in the literature. Rather, we use MCV to assess if there is a relation between the effects of various variables on tests, as is frequently done (e.g., Kan et al., 2013; Flynn et al., 2014).

So, we computed standardized SIRE differences for each test as described below. We also ran the admixture regression analyses for the genetic, brain volume, and admixture variables with cognitive test scores as the dependent variables. The effects on the eleven cognitive tests formed the vectors. These vectors were then correlated with and without corrections for test reliability. A genetic hypothesis for SIRE and ancestry differences predicts a positive manifold of genetic, brain volume, and SIRE/ancestry effects. The reason is that if genes related to *g* and brain volume are largely causative of differences both within and between groups, then these genes will cause similar patterns of effects across cognitive tests (te Nijenhuis et al., 2019). For example, if eduPGS effects are more pronounced on Matrix Reasoning than on Card Sorting, then we would expect ancestry effects also to be more pronounced if these ancestry effects on intelligence are due to eduPGS differences.

For these analyses, we computed the variables in such a way that they were conceptually and statistically independent:

1. Cognitive test *g* loadings were based on factor analysis using the Schmid-Leiman transformation. The Schmid-Leiman transformation separates *g* from non-*g* variance. Detailed results are provided in Supplementary File 1, Table S11.
2. The effects of MTAG-eduPGS, eduPGS and poly-eduPGS on cognitive test scores were the betas for the eduPGS predictor from the mixed-effects models. The mixed-effects models also included the cubic spline of age, sex, and the first twenty genetic PCs in addition to site and family fixed effects.
3. The effect of European admixture on cognitive test scores were the betas for the European ancestry predictor from the mixed-effects models which also include the cubic spline of age, sex, US born status, immigrant family status, SIRE fraction, and Hispanic ethnicity in addition to site and family fixed effects.
4. The effects of brain volume on cognitive test scores were the betas for the brain volume predictor from the mixed-effects models which also included the cubic spline of age, sex, height, US born status, immigrant family status, SIRE fraction, Hispanic ethnicity, and the first twenty genetic PCs in addition to site and family fixed effects.
5. Heritability estimates for the full sample were from the biometric analyses.
6. Black-White and Hispanic-White mean differences are computed as the sex- and age- corrected mean differences divided by the pooled standard deviations of both groups.

This procedure produced eight variables. These vectors were corrected for cognitive test reliabilities, based on validation, standardization, and other samples. See Supplementary File 1, Table S12a for the list of sources. Note that no reliability data were available for the Little Man Test. For this reason, we ran the analyses for both vectors uncorrected for reliability ($N = 11$) and for vectors corrected for reliability ($N = 10$). The uncorrected and corrected vectors were then correlated using Pearson’s correlation (with results based on Spearman’s correlation reported in Supplementary File 1, Table S12c).

A possible explanation for the association between the vectors of eduPGS-loadings and both the vectors of ancestry effects and SIRE differences is that these variables are all related to *g* loadings. According to this argument, these relations merely indicate that the relevant variables are related to phenotypic *g* (Kan, 2012; Chapter 4). To test this conjecture, we use partial correlations and control alternatively for the vector of *g* loadings and the vector of eduPGS effects.

2.3.7.2 Factor analysis

As an alternative analysis, we factor analyzed the eight vectors using principal factor analysis. This was done to determine if the SIRE/ancestry variables loaded strongly on the same factor that genetic variables did.

2.3.7.3 Relation between effects of eduPGS on test scores within SIRE groups and phenotypic differences between SIRE groups

We further examined the relation between the mean differences on tests *between* SIRE groups and the effects of eduPGS on tests *within* SIRE groups. We computed the betas for MTAG- and Poly-eduPGS for each SIRE group separately. We used mixed-effects models, which included the cubic spline of age, sex, and the first twenty genetic PCs in addition to site and family fixed effects, as above. We averaged the eduPGS effects for White and the respective non-White groups and then correlated the vector of average eduPGS effects with the standardized differences in cognitive ability.

2.4 Summary showing how statistical analyses are related

We carefully investigated a collection of closely interrelated hypotheses with seven sets of statistical analyses and then statistically connected the results using two sets of statistical analyses. As some readers may not be able to see the forest for the trees due to the large number of statistical analyses, we make explicit how the statistical analyses are related to each other.

Set 1: We carefully investigated cognitive differences, using multi-group confirmatory factor analysis, and assessed the extent to which these differences are measurement invariant and on *g*.
Set 2: We assessed the environmental and genetic influences on *g* and brain volume within groups. We incorporated the subtest heritability estimates into the final analysis using MCV.
Set 3: We ran global admixture analyses for *g* and brain volume. We explored these results carefully because the subsequent path, mediation, and MCV analyses depended on them.
Set 4 and Set 5: We carefully examined to what extent eduPGS/brain volume predict *g* within SIRE groups and within families.
Set 6: We incorporated our *g*, brain volume, and eduPGS variables in path and mediation analyses.
Set 7: We incorporated our *g*, brain volume, and eduPGS variables along with heritability estimates in MCV analyses.

2.5 Supplementary analyses

We included two sets of supplementary analyses. The first provides additional information regarding possible PGS bias across ancestries. The second provides additional information regarding the explanatory power of MRI-based brain differences. It made more sense to report these analyses separately than under the heading of any of the hypotheses listed above.

2.5.1 Group-level analyses for PGS and cognitive ability

Population-level analyses have frequently been employed to assess the cross-population validity of PGS (e.g., Chande et al., 2020; Martin et al., 2017). It is sometimes found that PGS differences are inconsistent with observed trait differences across populations. When this is the case, the cross-population validity of the PGS is called into question (e.g., Martin et al., 2017). So, we examined the correlation between several educational and intelligence PGS, on the one hand, and SIRE and national mean cognitive scores, on the other hand. High positive correlations between PGS differences and trait differences can increase our confidence that PGS differences correspond with causally-relevant genetic differences (Warne, 2021).

To maximize the number of groups, we used the dataset with all individuals, even those with missing data on admixture and *g*. We then decomposed the sample in 30 ethnocultural groups based on SIRE and nationality variables. For comparison, we alternatively decomposed the sample by the child's religion, which can overlap with ancestry (e.g., most Hindus are Asian Indians). We then correlated mean PGS and NIHTBX scores. We used seven different PGSs, based on different methods (GWAS, MTAG, and functional mapping) and on different criteria (education, cognitive ability/intelligence, and mathematical skills) because it is plausible that scores will vary by groups due to PGS construction and criteria. We focused on lead or functionally-informed SNPs because these are generally less influenced by confounding related to differences in linkage disequilibrium. The PGS used are as follows:

1. MTAG-eduPGS – These PGSs were used in the main analysis and are detailed above.
2. Poly-eduPGS - These PGSs were used in the main analysis and are detailed above.
3. Lee_GWASlead-eduPGS - Scores computed from the lead GWAS educational SNPs reported by Lee et al. (2018) ($N = 9{,}336$ SNPs were in the ABCD sample).
4. MTAG-cognitivePGS - Scores computed from the MTAG for cognitive ability SNPs from Lee et al. (2018) ($N = 8{,}898$ SNPs in the ABCD sample). These were constructed the same way as the MTAG-eduPGS, except that cognitive ability, not educational attainment, was the criterion.
5. MTAG-mathclassPGS - Scores computed from the MTAG for highest math class taken from Lee et al. (2018) ($N = 8{,}898$ SNPs in the ABCD sample). These were constructed the same way as the MTAG-eduPGS, except that highest math class taken, not educational attainment, was the criterion.
6. MTAG-mathabilityPGS - Scores computed from the MTAG for self-reported math ability SNPs from Lee et al. (2018) ($N = 8{,}898$ SNPs in the ABCD sample). These were constructed the same way as the MTAG-eduPGS, except that self-reported math ability, not educational attainment, was the criterion.
7. Hill_MTAG-IQPGS - Scores based on the lead independent SNPs identified using FUMA (functional mapping and annotation of genetic associations) from Hill et al. (2019) ($N = 512$ variants in the ABCD sample).

We correlated the seven PGS and NIHTBX scores across groups. When doing so, we weighted the groups by the square-root of the sample size, so that effects of larger sized groups did not dominate the results. Next, we output regression plots for PGS and NIHTBX. Details about the construction of the SIRE/national groups are provided in Supplemental File 2, Tab 1.

2.5.2 Multi-model MRI-based brain predictor, cognitive ability, and ancestry

In the main analyses, we focus on brain volume as a mediator of the relation between genetic ancestry and cognitive ability. It is possible that while brain volume could mediate the relation between European ancestry and ability, other brain variables could moderate this relation. So, to see how robust the results are, we run analyses using a multi-model MRI-based predictor of intelligence.

This variable, derived by applying machine learning to over 50,000 neurological variables, including both structural and functional MRI data, is detailed by Kirkegaard & Fuerst (2023). The brain predictor was trained to predict cognitive ability in only the non-Hispanic White sample to avoid ancestry confounding. In these supplementary analyses, we examine the relation between the MRI-based brain predictor of intelligence and genetic ancestry in the subsample of $N = 7464$ who had both MGCFA-based g scores and MRI-based brain predictor scores. For these analyses g, the brain predictor, and height were re-standardized in the subsample of $N = 7464$, so that the means were zero and the standard deviations were 1. We repeated the same analyses described in Section 2.2.6.6, replacing brain volume with the brain predictor variable.

2.6 Code and data

The R-code used for these analyses is provided in Supplemental File 1. Supplemental files and appendices can be found at: https://osf.io/wudtq/ The complete data set is available to qualified researchers at: https://nda.nih.gov/abcd

3. Results

3.1 Measurement of group differences

Descriptive statistics for the total sample and the four SIRE subsamples are shown in Table 1, along with the Cohen's d values between the White and the non-White groups. Based on conventional standards (Cohen, 1988), there are medium to large differences between Whites and non-White groups in g, eduPGS, and adjusted brain/intracranial volume. The difference in mean g between Whites and Blacks was typically sized at $d = 1.02$. However, the difference in mean g between Whites and Hispanics was smaller than typically found at $d = 0.38$. For context, C. Murray (2021) estimated the Black-White and Hispanic-White Full-Scale IQ differences to be $d = .85$ and $d = 0.62$, respectively, based on four recent IQ standardizations since the 2000s. Regarding brain volume, Jensen (1998, *p*. 438) notes a Black-White difference of $d = .76$ to .78 in brain weight based on autopsy data of individuals without brain pathologies. Also, Rushton & Ankney (2009) note a 63 and 32 gram Black-White and Hispanic-White difference, respectively, in brain weight; since standard deviations were not provided we were unable to convert these latter brain weight differences into standardized units. For comparison, the Black-White adjusted brain volume difference in this sample of healthy adolescents is approximately one standard deviation ($d = 1.12$).

We find substantially larger SES gaps than reported by Warne (2021). Whereas Warne (2021) reports a Black-White composite score SES gap — based on parental income, occupation, and educational level – of $d = 0.66$, we find a general factor of SES gap of $d = 1.82$. Our SES variable, however, was computed using a larger, more diverse set of variables and also using factor analysis. As a result, the values we report are not directly comparable to those reported by Warne (2021). In Supplementary File 2, we additionally report means, standard deviations, and Cohen's *d*s for the individual SES variables. This way interested readers can compare the SES-related differences in this dataset to those in other datasets. It can be seen, for example, that the average Black-White parental education and family income gaps are $d = 1.14$ and $d = 1.18$, respectively. So, in this sample, the Black-White educational and income differences are larger than those reported by Warne (2021).

Regarding Spearman's hypothesis, Appendix A, Table A2 shows the MGCFA model fit statistics for SIRE groups. The best fitting models were the model for the strong version of Spearman's hypothesis (M6A), in which g alone explained group differences, and the model for the weak version of Spearman's hypothesis in which the *g* factor explained most of the differences but the memory and executive function group factors also contributed to group differences (M6B).

Table 1. Descriptive statistics for the total sample and the four SIRE subsamples

	Full Sample $M \pm SD$	White $M \pm SD$	Hispanic $M \pm SD$	Black $M \pm SD$	Other $M \pm SD$	W-H *d*	W-B *d*	W-O *d*
Age	9.91 ± 0.62	9.93 ± 0.63	9.88 ± 0.63	9.91 ± 0.61	9.89 ± 0.62	0.08	0.03	0.06
European ancestry	0.75 ± 0.33	0.98 ± 0.05	0.60 ± 0.21	0.16 ± 0.11	0.62 ± 0.25	3.32	12.09	3.75
African ancestry	0.18 ± 0.31	0.01 ± 0.02	0.10 ± 0.14	0.82 ± 0.11	0.32 ± 0.26	-1.24	-14.88	-3.48
Amerindian ancestry	0.06 ± 0.14	0.01 ± 0.03	0.28 ± 0.19	0.01 ± 0.02	0.04 ± 0.09	-2.72	0.00	-0.73
South Asian ancestry	0.00 ± 0.02	0.00 ± 0.01	0.01 ± 0.01	0.00 ± 0.01	0.01 ± 0.05	-1.00	0.00	-0.52
East Asian ancestry	0.01 ± 0.03	0.00 ± 0.02	0.01 ± 0.02	0.01 ± 0.02	0.01 ± 0.07	-0.50	-0.50	-0.33
Frac White SIRE	0.73 ± 0.43	1.00 ± 0.03	0.67 ± 0.45	0.00 ± 0.00	0.39 ± 0.25	1.44	37.72	6.90
Frac Black SIRE	0.19 ± 0.38	0.00 ± 0.00	0.07 ± 0.23	1.00 ± 0.04	0.29 ± 0.26	-0.60	-53.43	-3.33
Frac. Native Amer. SIRE	0.02 ± 0.11	0.00 ± 0.00	0.03 ± 0.13	0.00 ± 0.00	0.18 ± 0.28	-0.46	NA	-1.92
frac_NOC SIRE	0.06 ± 0.23	0.00 ± 0.03	0.23 ± 0.42	0.00 ± 0.04	0.13 ± 0.34	-1.08	0.00	-1.11
Hispanic	0.20 ± 0.40	0.00 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	NA	NA	NA
g	0.00 ± 1.00	0.24 ± 0.86	-0.10 ± 0.99	-0.69 ± 1.07	-0.08 ± 1.07	0.38	1.02	0.36
SES	0.00 ± 1.00	0.45 ± 0.75	-0.38 ± 0.90	-0.98 ± 0.91	-0.39 ± 0.99	1.05	1.82	1.08
MTAG-eduPGS	0.00 ± 1.00	0.49 ± 0.77	-0.21 ± 0.75	-1.34 ± 0.57	-0.36 ± 0.86	0.92	2.50	1.09
Poly-eduPGS	0.00 ± 1.00	0.48 ± 0.78	-0.21 ± 0.78	-1.30 ± 0.60	-0.36 ± 0.85	0.88	2.39	1.07
Brain vol.	0.00 ± 1.00	0.26 ± 0.94	-0.18 ± 0.95	-0.64 ± 0.90	-0.13 ± 1.00	0.47	0.97	0.41
Brain vol. adj.	0.00 ± 1.00	0.29 ± 0.92	-0.21 ± 0.92	-0.74 ± 0.91	-0.13 ± 0.99	0.54	1.12	0.45
Intracranial vol.	0.00 ± 1.00	0.20 ± 0.96	-0.19 ± 1.01	-0.40 ± 0.94	-0.21 ± 1.01	0.40	0.63	0.42
Intracranial vol. adj.	0.00 ± 1.00	0.23 ± 0.94	-0.21 ± 1.01	-0.47 ± 0.95	-0.22 ± 1.00	0.46	0.74	0.48
Height	0.00 ± 1.00	-0.04 ± 0.97	-0.12 ± 0.97	0.26 ± 1.09	0.05 ± 0.97	0.08	-0.30	-0.09
Child US born	0.98 ± 0.15	0.99 ± 0.11	0.94 ± 0.24	0.98 ± 0.15	0.98 ± 0.14	0.32	0.08	0.09
Immigrant family	0.28 ± 0.45	0.18 ± 0.38	0.73 ± 0.44	0.14 ± 0.35	0.20 ± 0.40	-1.39	0.11	-0.05
N	10,245	5,858	2,005	1,642	740			

Notes: W-H *d*, W-B *d*, and W-O *d* are the Cohen's *d* between, respectively, the White and Hispanic. White and Black, and White and Other SIRE groups. A positive value indicates an advantage for the White group.

3.2 Measurement invariance across genetic ancestry

Using the same three-broad-factor model to test for strict measurement invariance between SIRE groups, we applied LSEM to the full sample ($N = 10,245$) using 11 focal points (representing 10% ancestry increments, ranging from 0% to 100% European ancestry). Across the full range of European ancestry, there was no deviation from strict MI. These results thus agree with those of Lasker et al. (2019), who found that strict MI held across European genetic ancestry in a large African and European American sample of adolescents, and Lasker (2019) who additionally reports results based on a smaller pediatric dataset including Blacks, Whites, and Hispanics. This indicates that the observed difference in *g* cannot be attributed to other causes that differ between races, such as stereotype threat (Lubke et al., 2003).

3.3 Within-group heritability and shared environmentality

Table 2 shows the number of DZ and MZ twin pairs along with the a'c'e' and ACE estimates for adjusted brain and intracranial volume. The a'c'e' estimates for the unadjusted variables are additionally provided in the Supplementary Material. For adjusted brain and intracranial volume, heritability estimates were high for all groups ($h^2 = .75$ to $.95$), while the shared environmental variance was low ($c^2 = .00$ to $.17$), consistent with previously reported values for this age group (Jansen et al., 2015). With regards to brain volume, heritability estimates were somewhat higher, while shared environmental estimates were lower for the non-White sample ($h^2 = .94$; $c^2 = .00$) than for the White sample ($h^2 = .79$; $c^2 = .11$), but this difference was not statistically significant. For intracranial volume, heritability and shared environment were very similar for the combined non-White sample ($h^2 = .88$; $c^2 = .07$) and for the White sample ($h^2 = .83$; $c^2 = .10$).

Table 2. Variance component estimates for brain volume and intracranial volume

	N_{DZpairs}	N_{MZpairs}	a'	*SE*	c'	*SE*	e'	*SE*	a^2	c^2	e^2
Brain Volume											
All available	443	333	.87	.04	.42	.08	.28	.01	.75	.17	.08
White-Hisp.-Black-Other	443	333	.87	.04	.42	.08	.28	.01	.75	.17	.08
White	426	323	.87	.04	.41	.08	.27	.01	.76	.16	.08
Hispanic-Black – Other	280	214	.89	.05	.34	.12	.30	.02	.79	.11	.09
Hispanic	146	109	.97	.00	.00	NA	.24	.02	.94	.00	.06
Black	43	34	.98	.01	.00	.00	.21	.03	.95	.00	.04
Other	84	56	.93	.08	.27	.29	.24	.03	.87	.07	.06
Intracranial Volume											
All available	443	333	.89	.04	.39	.08	.24	.01	.79	.15	.06
White-Hisp.-Black-Other	426	323	.89	.04	.39	.08	.25	.01	.78	.15	.06
White	280	214	.91	.05	.32	.13	.26	.01	.83	.10	.07
Hispanic-Black – Other	146	109	.94	.06	.26	.23	.24	.02	.88	.07	.06
Hispanic	43	34	.95	.11	.24	.45	.21	.03	.90	.06	.04
Black	84	56	.90	.08	.39	.19	.21	.02	.81	.15	.05
Other	19	19	.93	.03	.00	.00	.36	.07	.87	.00	.13

Notes: a and a^2 represent, respectively, the genetic-phenotypic correlations and proportion of variance due to additive genetic effects; c and c^2 represents, respectively, the shared environment-phenotypic correlations and proportion of variance due to shared environmental effects; e and e^2 represents, respectively, the non-shared environment-phenotypic correlations and proportion of variance due to non-shared environmental effects; $N_{\text{pairs DZ}}$ is the number of dizygotic twin pairs; $N_{\text{pairs MZ}}$ is the number of monozygotic twin pairs. *SE* = standard error.

Table 3 shows the ACE estimates for the NIHTBX and *g* scores. More detailed results are provided in Supplementary File 1. Consistent with previously reported results (Pesta et al., 2020, 2023), the heritability estimates of the NIHTBX scores were moderate-to-high for the total, the White, and the combined non-White samples (h^2 = .56 to .64). In line with this, the shared environmental variances were low (c^2 = .00 to .20). For NIHTBX, the heritability was somewhat higher, and shared environmentality lower for the White sample ($h^2 = .64$; $c^2 = .00$) than the combined non-White sample ($h^2 = .56$; $c^2 = .20$), though these differences were not statistically significant. Regarding *g* scores, the heritability was also somewhat higher, and shared environmentality lower for the White sample ($h^2 = .67$; $c^2 = .00$) than the non-White sample ($h^2 = .59$; $c^2 = .16$), though, again, these differences were not statistically significant. For comparison, in their meta-analysis, Pesta et al. (2020; Table 4) found that heritabilities and shared environmentalities were approximately the same among Hispanics, Blacks, and Whites.

Supplementary File 1, Table S21, additionally reports the heritabilities for each of the 11 subtests used to create our *g* scores. These subtest heritabilities were used in the MCV analysis.

Table 3. Variance component estimates for NIH Toolbox cognitive scores

	$N_{DZpairs}$	$N_{MZpairs}$	a’	*SE*	c’	*SE*	e’	*SE*	a^2	c^2	e^2
NIHTBX											
All available	443	332	.76	.06	.35	.11	.55	.02	.57	.13	.30
White-Hisp.-Black –Other	441	330	.76	.06	.35	.11	.55	.02	.57	.12	.30
White	291	219	.80	.02	.00	.00	.60	.03	.64	.00	.36
Hispanic-Black – Other	150	111	.75	.09	.45	.14	.50	.04	.56	.20	.25
Hispanic	43	34	.82	.14	.40	.28	.41	.06	.67	.16	.17
Black	87	58	.73	.13	.40	.22	.55	.05	.54	.16	.30
Other	20	19	.76	.27	.32	.58	.56	.10	.58	.10	.32
G											
All available	434	319	.78	.05	.36	.11	.52	.02	.60	.13	.27
White-Hisp.-Black-Other	430	319	.78	.05	.35	.11	.52	.02	.61	.12	.27
White	291	220	.82	.02	.00	.00	.57	.03	.67	.00	.33
Hispanic-Black – Other	139	99	.77	.09	.41	.16	.50	.04	.59	.16	.25
Hispanic	43	34	.32	.25	.82	.08	.48	.06	.10	.67	.23
Black	87	56	.80	.12	.30	.29	.52	.05	.64	.09	.27
Other	9	9	.80	.11	.00	NA	.60	.15	.64	.00	.36

Notes: a and a^2 represents, respectively, the genetic-phenotypic correlations and proportion of variance due to additive genetic effects; c and c^2 represent, respectively, the shared environment-phenotype correlations and proportion of variance due to shared environmental effects; e and e^2 represent, respectively, the non-shared environment-phenotype correlations and proportion of variance due to non-shared environmental effects; $N_{pairs\ DZ}$ is the number of dizygotic twin pairs; $N_{pairs\ MZ}$ is the number of monozygotic twin pairs. *SE* = standard error.

We provide more details on the low heritability of *g* scores for the Hispanic group. We found that ACE estimates for *g* were unstable in this small twin sample. The SEM model had difficulty distinguishing between the A and the C component, which was not surprising because there were only 43 DZ twins and 34 MZ twins. When Black and Other twins were added, the ACE estimates became both moderate-to-high and stable. The discrepancy between the ACE estimates for *g* and for NIHTBX, among Hispanics, resulted from both a difference in scores and a difference in twin pairs (while the numbers were the same, the sets were not, owing to missing data in both variables). When we computed the ACE estimates for each subtest, we found that the averages were similar for Hispanics (A = .28, C = .17, E = .54) and for Whites (A = .33, C = .04, E = .63). The ACE estimates by test are provided in Supplementary File 2. Generally, ACE estimates based on small samples can be very unstable, which is why we limited our research focus to estimates for the White and combined Hispanic-Black-Other groups.

3.4 Admixture regression for g

The triangle admixture plot for the full sample is shown in Figure 2. Hispanics had high variability along the Amerindian-European axis and moderate variability along the African-European axis. The African American group and the Other group had, respectively, moderate and high variability along the African-European axis. White Americans exhibited little variability in ancestry. This pattern of admixture is typical for the US (Bryc et al., 2015).

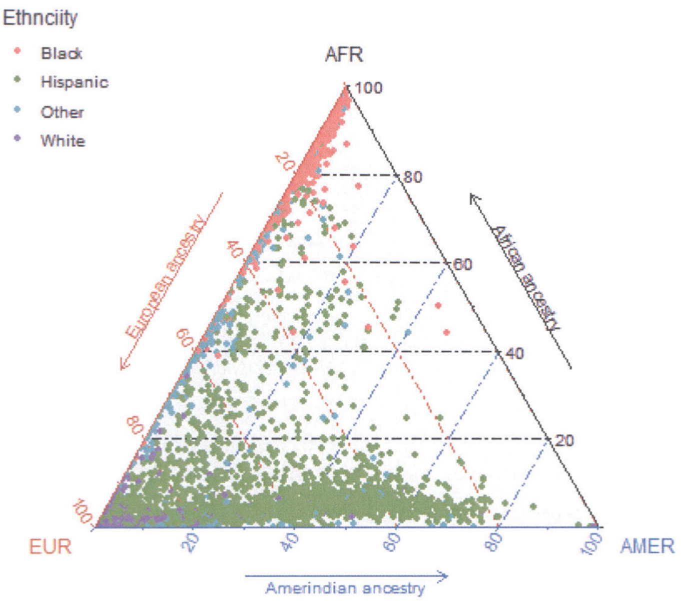


Figure 2. Admixture triangle plot for SIRE groups in the ABCD sample

The correlation matrices for the full sample and the SIRE subsamples are provided in Supplemental File 1 (Tables S3a-S3e). Table 4 shows the results of the admixture regression analyses for *g*.

Table 4. Regression results for the effect of genetic ancestry on *g* in the full sample

	Model 1		Model 2a		Model 2b		Model 3a		Model 3b	
	b	*p*	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.75 (0.38)	0.051	0.62 (0.38)	0.100	0.43 (0.37)	0.247	-1.12 (0.18)	**<0.001**	-0.76 (0.17)	**<0.001**
frac_Black SIRE	-0.99 (0.03)	**<0.001**	0.03 (0.09)	0.733	0.06 (0.08)	0.501	0.04 (0.07)	0.566	0.01 (0.06)	0.825
frac_Native Amer. SIRE	-0.36 (0.09)	**<0.001**	-0.07 (0.09)	0.463	0.02 (0.09)	0.791	-0.06 (0.09)	0.533	0.03 (0.09)	0.717
frac_NOC SIRE	-0.42 (0.05)	**<0.001**	-0.09 (0.05)	0.079	-0.01 (0.05)	0.760	-0.08 (0.05)	0.102	-0.01 (0.05)	0.846
Hispanic	-0.30 (0.03)	**<0.001**	0.01 (0.04)	0.791	0.06 (0.04)	0.135	-0.04 (0.03)	0.223	0.03 (0.03)	0.336
Child US Born	0.12 (0.06)	**0.049**	0.14 (0.06)	**0.027**	0.18 (0.06)	**0.002**	0.12 (0.06)	**0.046**	0.18 (0.06)	**0.003**
Immigrant Family	0.10 (0.03)	**<0.001**	0.15 (0.03)	**<0.001**	0.08 (0.03)	**0.001**	0.16 (0.03)	**<0.001**	0.09 (0.02)	**<0.001**
sex [M]	-0.01 (0.02)	0.483	-0.01 (0.02)	0.696	-0.01 (0.02)	0.468	-0.01 (0.02)	0.721	-0.01 (0.02)	0.491
Age	-0.06 (0.04)	0.118	-0.05 (0.04)	0.233	-0.05 (0.04)	0.234				
Age'	0.08 (0.05)	0.120	0.06 (0.05)	0.222	0.06 (0.05)	0.247				
African ancestry			-1.34 (0.11)	**<0.001**	-0.87 (0.11)	**<0.001**				
Amerindian ancestry			-1.56 (0.12)	**<0.001**	-0.90 (0.11)	**<0.001**				
East Asian ancestry			0.13 (0.33)	0.703	0.21 (0.32)	0.516				
South Asian ancestry			0.42 (0.56)	0.450	0.76 (0.54)	0.161				
SES					0.33 (0.01)	**<0.001**			0.33 (0.01)	**<0.001**
European ancestry							1.35 (0.08)	**<0.001**	0.81 (0.08)	**<0.001**
Age							-0.00 (0.01)	0.891	-0.00 (0.01)	0.767

Random effects					
σ^2	0.41	0.41	0.41	0.41	0.41
τ_{00}	0.46 site_id_l:rel_family_id 0.02 site_id_l	0.44 site_id_l:rel_family_id 0.03 site_id_l	0.36 site_id_l:rel_family_id 0.05 site_id_l	0.44 site_id_l:rel_family_id 0.03 site_id_l	0.37 site_id_l:rel_family_id 0.05 site_id_l
ICC	0.54	0.53	0.50	0.53	0.50
N	22 site_id_l 8600 rel_family_id	22 site_id_l 8600 rel_family_id	22 site_id_l 8600 rel_family_id	22 site_id_l 8600 rel_family_id	22 site_id_l 8600 rel_family_id
Observations	10245	10245	10245	10245	10245
Marginal R^2 / conditional R^2	0.145 / 0.606	0.178 / 0.615	0.247 / 0.625	0.175 / 0.615	0.246 / 0.625

Notes: Unstandardized beta coefficients (*b*) and *p*-values (*p*) from the mixed-effects models, with recruitment site and family common factors treated as random effects are shown. The values in parentheses are standard errors. The marginal and conditional R^2 of the mixed effects model are shown at the bottom. ICC = intraclass correlation coefficient. Note, only linear age was used in context to European ancestry for reasons discussed in the Methods section.

In this table, Model 1 shows the results with SIRE but not genetic ancestry variables, while Model 2a adds the four non-European genetic ancestries. Model 2b adds SES to Model 2a. Models 3a and 3b repeat Models 2a and 2b, replacing the four non-European ancestries with European ancestry. In these last two Models (3a & 3b), age, instead of cubic spline of age is used. This was done so that effect plots could be reported for the results.

As seen in Model 1 and 2, SIRE variables are unrelated to *g*, independent of genetic ancestry. Additionally, African, Amerindian, and European ancestry had large effects on *g* even when the general factor of SES was added to the model.

The effects of European, African, and Amerindian ancestry on *g*, stratified by SIRE, are summarized in Table 5. The betas are from Models 1b to 2b shown in Tables S4a to S4e. Statistically significant ($p < .05$) values are shown in bold. The complete model results are provided in Supplementary File 1. The relation between European, African, and Amerindian ancestry and *g* is robust to SIRE stratification.

Table 5. Validities (*b*) from the mixed-effects regression models for genetic ancestry and *g* in the full sample (All) and in the SIRE subsamples

Predictor	Criterion	Controls	All	White	Hispanic	Black	Other
European ancestry	*g*		**1.35**	**1.33**	**1.33**	**1.06**	**0.96**
	g	**w/ SES**	**0.81**	**0.79**	**0.82**	**0.84**	**0.54**
African ancestry	*g*		**-1.34**	**-1.72**	**-1.10**	**-1.10**	**-1.21**
	g	**w/ SES**	**-0.87**	-0.81	**-0.65**	**-0.88**	**-0.79**
Amerindian ancestry	*g*		**-1.56**	**-1.86**	**-1.40**	**-3.84**	**-1.32**
	g	**w/ SES**	**-0.90**	**-1.11**	**-0.86**	**-3.12**	-0.79
N			10245	5858	2005	1642	740

Note: Statistically significant results ($p < .05$) are presented in bold.

The partial residual plot with respect to European ancestry and *g* is depicted in Figure 3. This plot shows the effect of European ancestry on *g* holding everything else in the regression model constant. Supplementary File 1, Figures S4a-S4e provide the residual plots for each SIRE subsample. This linear relationship is evident in each SIRE group and in the full sample.

Additionally, Table 6 reports the beta coefficients for the sibling and the comparable full sample analyses with either *g* or NIHTBX as the dependent variable. The complete model results are provided in Supplementary File 1, Table S15c. Within sibships, European ancestry was not significantly related to intelligence, though the standard errors were large. The shrinkage of the within-sibship coefficients, relative to the full sample ones, was 33%. Generally, as Conley &

Fletcher (2017) note, within-sibship ancestry analyses are highly underpowered and, as such, unreliable.

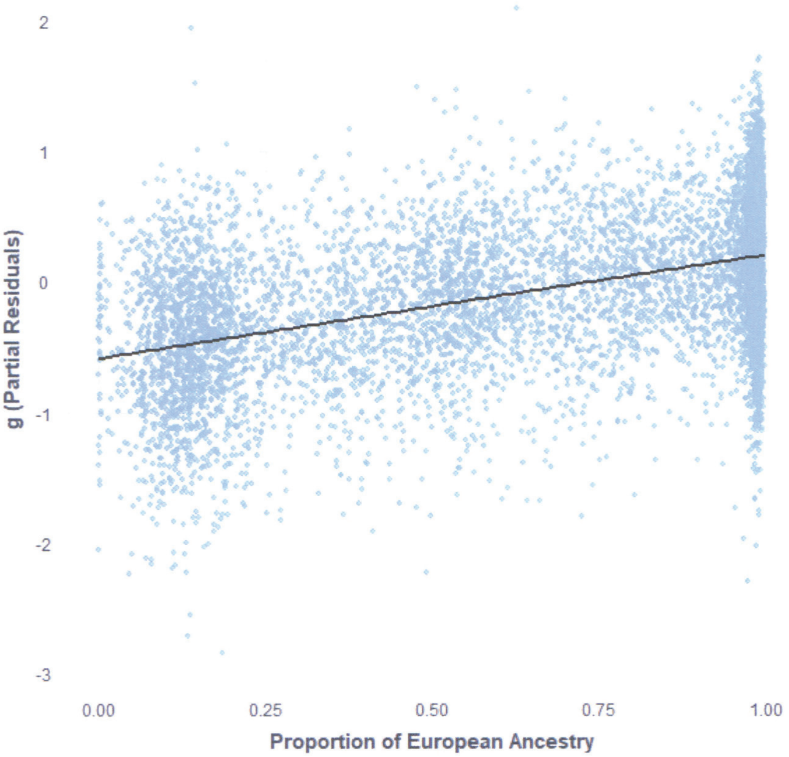


Figure 3. Partial residuals and estimated regression fits for European ancestry in the admixture regression with *g* as the dependent variable

Table 6. Validities (*b*) for the within- and between-sibship and full sample analyses with *g* or NIHTBX as the dependent variable and European ancestry as the predictor. Controls include sex, age, East Asian and South Asian ancestry

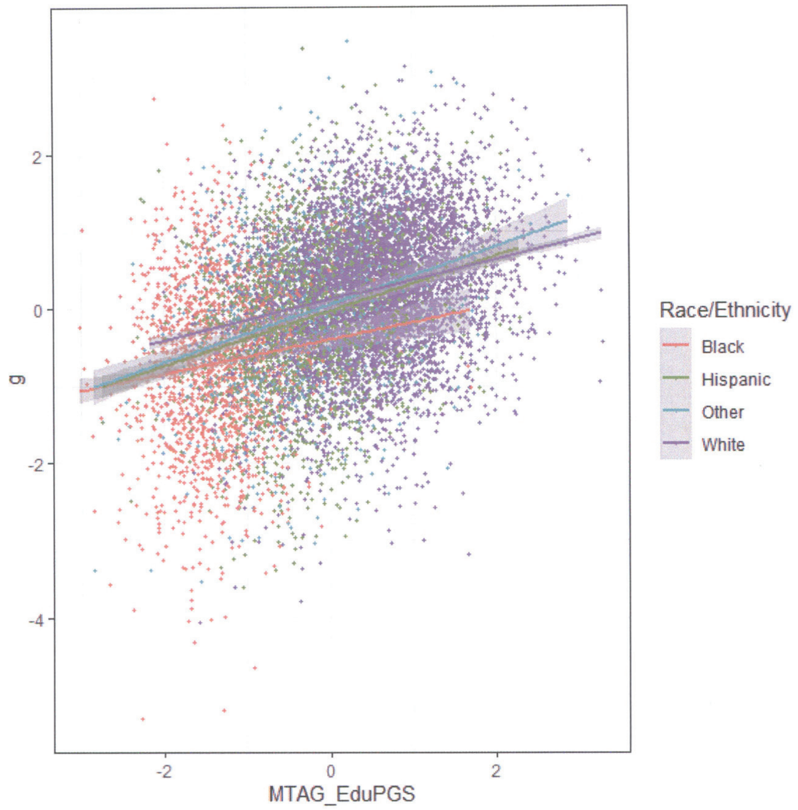
			b (*SE*, *N*)		
Predictor	Criterion	Within sibships	Between sibships	Full sample	Shrinkage
European	*g*	0.62	**1.02**	**0.93**	33%
Ancestry		(0.55; 2291)	(0.06; 2291)	(0.02; 10370)	
European	NIHTBX	0.83	**1.09**	**1.23**	33%
Ancestry		(0.53; 2222)	(0.06; 2222)	(0.03; 10040)	

Notes: Statistically significant results ($p < .05$) are presented in bold. Shrinkage % = 1 – (Within sibships *b* / Full sample *b*).

3.5 Polygenic scores

Next, we report the relation between eduPGS and *g*. For these analyses *g* is the dependent variable and the polygenic scores are the main predictors. The regression plots for eduPGS and *g* in the Black (red), Hispanic (green), Other (blue), and White (purple) samples are shown in Figure 4.

a. MTAG-eduPGS



b. Poly-eduPGS

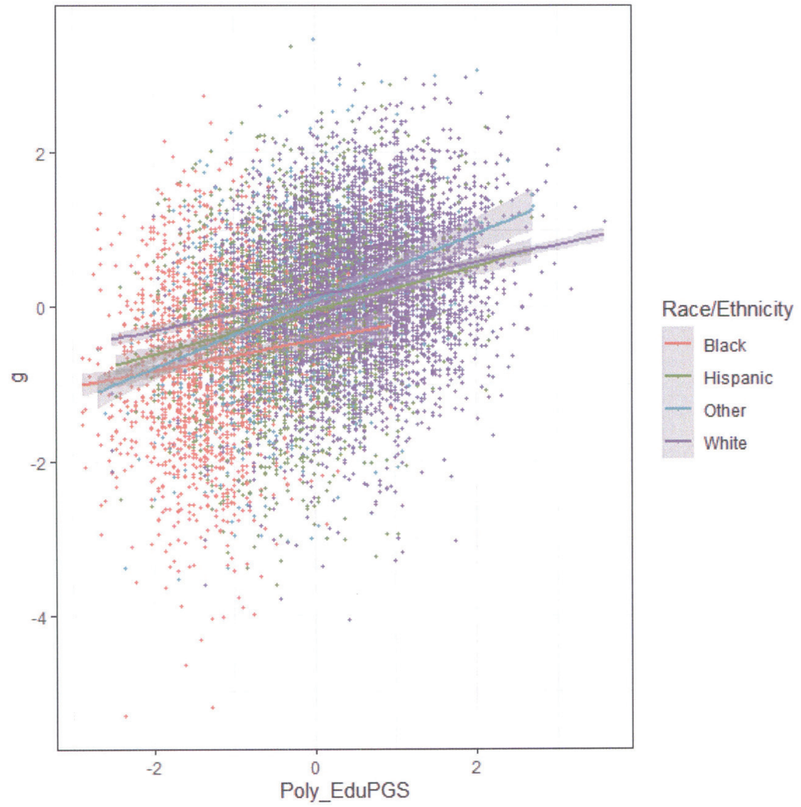


Figure 4. Regression plot for the predictive validity of eduPGS with respect to *g* in the Black (red), Hispanic (green), Other (blue), and White (purple) samples

We report results from the regression models for eduPGS and *g*. Tables S5a-S5e show the complete results. Table 7 summarizes the results from these analyses. eduPGS is statistically significantly related to *g* in all groups. The validities for Black Americans are reduced by around one-third relative to that for Whites. This finding is consistent with previously reported results (Lasker et al., 2019). It should be noted that the validities, even among Whites, are lower than those reported for European-ancestry discovery samples (e.g., Lee et al., 2019). This is theoretically expected because these validities concern out-of-sample predictions, which do not suffer from overfitting (Choi et al., 2020).

Table 7. Validities (*b*) from the mixed-effects regression models for eduPGS and *g* in the full sample and in the SIRE subsamples

Predictor	Criterion	Controls	All	White	Hispanic	Black	Other
MTAG-eduPGS	*g*	20PCs, sex, age	**0.26**	**0.26**	**0.26**	**0.16**	**0.32**
Poly-eduPGS	*g*	20PCs, sex, age	**0.23**	**0.22**	**0.20**	**0.16**	**0.42**
N			10245	5858	2005	1642	740

Notes: Statistically significant results ($p < .05$) are presented in bold.

Finally, Table 8 reports the beta coefficients for the sibling and the comparable full sample analyses with either *g* or NIHTBX as the dependent variable. The complete model results are provided in Supplementary File 1, Table S15b. Within sibships, both EduPGS were statistically significant predictors of intelligence. The shrinkage of the within-sibship coefficients, relative to the full sample ones, ranged from 20 to 27%. This is in line with the within-sibship shrinkage of 22% (95% CI: 6%, 37%) reported by Howe et al. (2022) for cognitive ability.

Table 8. Validities (*b*) for the within- and between-sibship and full sample analyses with *g* or NIHTBX as the dependent variable and EduPGS as the predictors. Controls include sex, age, and 20 principal components

		b (*SE*; *N*)			
Predictor	**Criterion**	**Within sibships**	**Between sibships**	**Full sample**	**Shrinkage**
MTAG- eduPGS	*g*	**0.20** (0.05; 2291)	**0.28** (0.03; 2291)	**0.25** (0.01; 10369)	20%
MTAG- eduPGS	NIHTBX	**0.21** (0.05; 2323)	**0.30** (0.03; 2323)	**0.27** (0.01; 10039)	22%
Poly-eduPGS	*g*	**0.17** (0.05; 2302)	**0.19** (0.03; 2302)	**0.22** (0.01; 10369)	23%
Poly-eduPGS	NIHTBX	**0.16** (0.05; 2323)	**0.22** (0.03; 2323)	**0.22** (0.01; 10039)	27%

Notes: Statistically significant results ($p < .05$) are presented in bold. Shrinkage % = 1 – (Within sibships *b* / Full sample *b*).

3.6 Brain volume

3.6.1 Brain volume and *g*

Table 9 summarizes the effects (*b*) for brain volume on *g* for the full sample and the SIRE subsamples. For these analyses, the adjusted brain volume variable is used. This adjusted variable was then included in the mixed-effects regression model along with the first twenty genetic PCs.

The complete results are provided in Supplementary Tables S8a-S8e. As can be seen, brain volume has a weak relation to *g* in the full sample and in all SIRE subsamples (*b*s = .12 to .25). Adjusting for genetic PCs attenuated the effect somewhat in the full sample and also in the three non-White samples; the effects remained statistically significant in all cases.

Table 9. Validities (*b*) from mixed-effects regression model for brain volume and *g* in the full sample and in the SIRE subsamples

Predictor	Criterion	Controls	All	White	Hispanic	Black	Other
Brain volume adj.	*g*		**0.22**	**0.13**	**0.18**	**0.15**	**0.25**
Brain volume adj.	*g*	20 PCs	**0.14**	**0.13**	**0.12**	**0.13**	**0.20**
N			10245	5858	2005	1642	740

Note: Statistically significant results ($p < .05$) are presented in bold.

Finally, Table 10 reports the beta coefficients for the sibling and the comparable full sample analyses with either *g* or NIHTBX as the dependent variable. The complete model results are provided in Supplementary File 1, Table S15a. Within sibships, brain volume was a statistically significant predictor of intelligence. The shrinkage of the within-sibship coefficients, relative to the full sample coefficients, ranged from 0 to 6%.

Table 10. Validities (*b*) for the within- and between-sibship and full sample analyses with *g* or NIHTBX as the dependent variable and brain volume as the predictor

Predictor	**Criterion**	**Controls**	**Within sibships**	**Between sibships**	**Full sample**	**Shrinkage**
Brain volume	*g*	20PCs, sex, age	**0.15** (0.04; 2258)	**0.16** (0.03; 2258)	**0.15** 0.01; 1024))	0%
Brain volume	NIHTBX	20PCs, sex, age	**0.15** (0.04; 2189)	**0.15** (0.03; 2189)	**0.16** (0.01; 9920)	6%

Notes: Statistically significant results ($p < .05$) are presented in bold. Shrinkage % = 1 – Within (Sibships *b* / Full Sample *b*). Cells show *b* (*SE*; *N*).

3.6.2 Brain volume, SIRE, and genetic ancestry

Table 11 shows the results from the admixture regression analysis for both brain volume and intracranial volume. In this table, Model 1a and 1b show the results for brain volume as the dependent variable, while Model 2a and 2b show those when intracranial volume is the dependent variable. In both Model 1b and 2b, European ancestry is used as the only ancestry predictor.

Table 11. Regression results for the effect of genetic ancestry on brain and intracranial volume in the full sample

	Brain volume predicted				Intracranial volume predicted			
	Model 1a		Model 1b		Model 2a		Model 2b	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.54 (0.33)	0.098	-0.26 (0.16)	0.095	0.25 (0.32)	0.437	-0.41 (0.17)	**0.016**
African ancestry	-0.89 (0.09)	**<0.001**			-0.68 (0.09)	**<0.001**		
Amerindian ancestry	-0.96 (0.10)	**<0.001**			-0.71 (0.09)	**<0.001**		
East Asian Ancestry	-0.39 (0.29)	0.175			-0.22 (0.27)	0.419		
South Asian Ancestry	-1.73 (0.48)	**<0.001**			-1.68 (0.45)	**<0.001**		
frac_Black SIRE	-0.00 (0.08)	0.969	0.01 (0.06)	0.820	-0.00 (0.07)	0.998	0.01 (0.05)	0.913
frac_Native American SIRE	0.05 (0.08)	0.551	0.04 (0.08)	0.598	0.00 (0.08)	0.999	-0.01 (0.07)	0.924
frac_NOC SIRE	0.03 (0.04)	0.464	0.03 (0.04)	0.456	0.01 (0.04)	0.799	0.01 (0.04)	0.792
Hispanic	-0.00 (0.03)	0.894	-0.01 (0.03)	0.627	0.01 (0.03)	0.692	0.01 (0.03)	0.761
Child US Born	-0.02 (0.05)	0.720	-0.02 (0.05)	0.679	-0.06 (0.05)	0.209	-0.06 (0.05)	0.192
Immigrant Family	0.03 (0.02)	0.161	0.03 (0.02)	0.184	0.02 (0.02)	0.329	0.02 (0.02)	0.372
Sex [M]	0.94 (0.02)	**<0.001**	0.94 (0.02)	**<0.001**	0.86 (0.01)	**<0.001**	0.86 (0.01)	**<0.001**
Age'	-0.08 (0.03)	**0.020**			-0.05 (0.03)	0.114		
Age"	-0.01 (0.05)	0.736			-0.00 (0.04)	0.950		
Height	0.16 (0.01)	**<0.001**	0.16 (0.01)	**<0.001**	0.19 (0.01)	**<0.001**	0.19 (0.01)	**<0.001**
SES	0.10 (0.01)	**<0.001**	0.11 (0.01)	**<0.001**	0.10 (0.01)	**<0.001**	0.10 (0.01)	**<0.001**
European ancestry			0.91 (0.07)	**<0.001**			0.68 (0.07)	**<0.001**
Age			-0.09 (0.01)	**<0.001**			-0.05 (0.01)	**<0.001**
Random effects								
σ^2	0.27		0.27		0.24		0.24	
τ_{00}	$0.36_{\text{site\_1\_id\^{}rel\_family\_id}}$ $0.01_{\text{site\_id\_1}}$		$0.36_{\text{site\_1\_id\^{}rel\_family\_id}}$ $0.01_{\text{site\_id\_1}}$		$0.30_{\text{site\_1\_id\^{}rel\_family\_id}}$ $0.17_{\text{site\_id\_1}}$		$0.30_{\text{site\_1\_id\^{}rel\_family\_id}}$ $0.17_{\text{site\_id\_1}}$	
ICC	0.58		0.58		0.66		0.66	
N	$22_{\text{site\_id\_1}}$ $8600_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $8600_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $8600_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $8600_{\text{rel\_family\_id}}$	
Observations	10245		10245		10245		10245	
Marginal R^2/ Conditional R^2	0.368/0.732		0.367/0.731		0.289/0.761		0.289/0.761	

Notes: Unstandardized beta coefficients (*b*) and *p*-values (*p*) from the mixed-effects models, with recruitment site and family common factors treated as random effects are shown. Values in parentheses are standard errors. The marginal and conditional R^2 of the mixed effects model are shown at the bottom. ICC = Intraclass correlation coefficient. Note, only linear age was used in context to European ancestry for reasons discussed in the Methods section.

As seen in all models, there was no association between SIRE and brain volume independent of genetic ancestry. Moreover, the effect of African ($b = -.89$) and Amerindian ($b = -.96$) ancestry, or conversely European ($b = .91$), on brain volume was largely independent of sex, age, height, migrant status, SIRE, the general factor of SES, and both site- and family-related fixed effects. For intracranial volume the ancestry effects were moderate: European ($b = .68$), African ($b = -.68$) and Amerindian ($b = -.71$).

The associations with European, African, and Amerindian genetic ancestry were stronger for brain volume than for intracranial volume. It is not clear why this is the case as these two variables correlate at $r = .86$ in the full sample. Both height and SES were statistically significantly related to brain and intracranial volume. These associations could be a result of genetic and environmental factors.

3.6.3 Brain/intracranial volume within SIRE groups

The effects of European, African, and Amerindian ancestry on brain and intracranial volume, stratified by SIRE, are summarized in Table 12. The betas are from Models 1a to 2b from Tables S6a to S6e, which report the equivalent model results from Table 11, above, for each SIRE group. Statistically significant ($p < .05$) values are shown in bold. The complete model results are provided in the Supplementary File 1. Generally, the finding of a positive effect of European ancestry and a negative effect of both African and Amerindian ancestry is robust to SIRE stratification.

The negative coefficients for African ancestry are very consistent with the anthropological record according to which sub-Saharan African populations have smaller cranial capacities than European populations (Clark & Henneberg, 2021; Coon, 1982; Hambly, 1947; Lewis et al., 2011; Smith & Beals, 1990). While some datasets also show that Amerindian populations tend to have smaller cranial capacities than European populations (e.g., Lewis et al., 2011), American Indians had cranial capacities very similar to Europeans in the largest dataset published. Specifically, Native Americans (excluding Inuit and Aleut) populations had an average cranial capacity of 1359 cm^3 ($k = 35$), as compared to 1365 cm^3 for Europeans ($k = 9$), and 1282 cm^3 for Sub-Saharan Africans (excluding Khoisan) ($k = 6$) (Smith & Beals, 1990). See, similarly: Hambly (1947). It is not clear why Amerindian ancestry is strongly negatively related to MRI-measured intracranial volume given the data reported by Smith and Beals (1990).

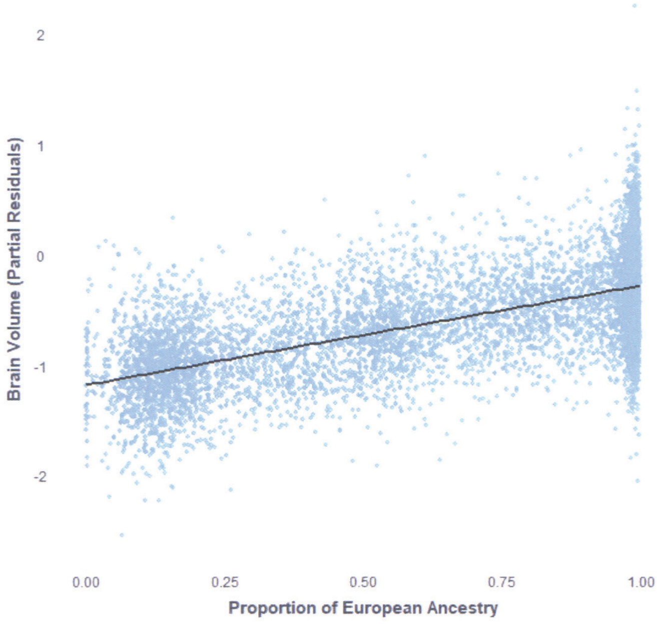
Table 12. Validities (*b*) from the mixed-effects regression models for genetic ancestry and brain and intracranial volume in the full sample and in the SIRE subsamples

Predictor	Criterion	All	White	Hispanic	Black	Other
European ancestry	Brain volume	**0.91**	**0.85**	**0.91**	**0.96**	**0.80**
African ancestry	Brain volume	**-0.89**	-0.74	**-0.73**	**-1.02**	**-0.84**
Amerindian ancestry	Brain volume	**-0.96**	**-1.02**	**-0.92**	**-2.80**	**-0.91**
European ancestry	Intracranial volume	**0.68**	**0.65**	**0.61**	**0.79**	**0.50**
African ancestry	Intracranial volume	**-0.68**	-0.40	**-0.54**	**-0.84**	**-0.52**
Amerindian ancestry	Intracranial volume	**-0.71**	**-0.93**	**-0.59**	**-2.45**	-0.42
N		10245	5858	2005	1642	740

Note: Statistically significant results ($p < .05$) are presented in bold.

Finally, Figure 5 depicts the partial residual plots with respect to European ancestry and both brain and intracranial volume.

a. Brain volume.



b. Intracranial volume

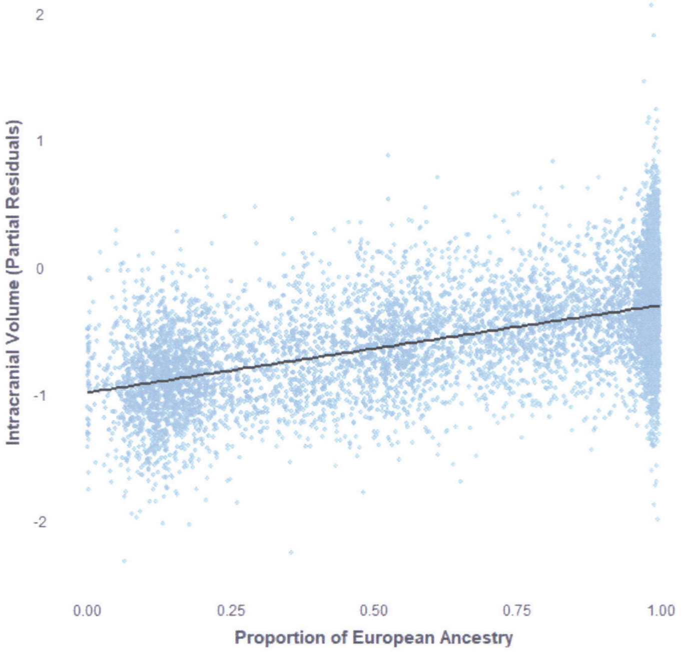


Figure 5. Partial residuals and estimated regression fits for European ancestry in the admixture regression with brain and intracranial volume as the dependent variables

As can be seen, there is a clear linear relationship between European ancestry and both brain and intracranial volume. Supplementary File 1, Figures S5a1-S53e2 provide the residual plots for each SIRE subsample. This linear relationship is evident within each SIRE group.

3.6.4 SES, brain volume, and *g* among adopted and biological children

The full results for the adopted and biological child analyses are provided in Supplementary File 1, Table S14. In the model for *g*, the betas for SES were $b = -.05$ (*S.E.*: 0.12) and .36 (*S.E.*: 0.02) for the adoptive and biological sample, respectively. In the model for brain volume, the betas for SES were $b = .01$ (*S.E.*: 0.10) and .11 (*S.E.*: 0.02) in the adoptive and biological sample, respectively. We reran the adoptive analyses, this time including an interaction term for adoptive status and SES. Adoption status significantly predicted *g* (*t*-value = -3.405), but not brain volume. Overall, these adoptive results suggest that the association between SES and *g*, in this ABCD cohort, may not be mediated by parental environment. This may also be the case for brain volume, but the power is too limited to draw conclusions.

It is also notable that child eduPGS predicted family SES about the same as child cognitive ability. The results are summarized in Table 13, with the full results reported in Table 5 of Supplemental File 2.

Table 13. Validities (*b*) from the mixed-effects regression models for eduPGS and SES in the full sample and in the SIRE subsamples

Predictor	Criterion	Controls	All	White	Hispanic	Black	Other
MTAG eduPGS	SES	20PCs	**0.19**	**0.21**	**0.15**	**0.14**	**0.12**
Poly eduPGS	SES	20PCs	**0.18**	**0.18**	**0.14**	**0.18**	**0.23**
N			10245	3654	1500	1130	499

Moreover, as seen in Table S15 of Supplemental File 2, the percentage with which each of the seven SES indicators mediated the relation between eduPGS and *g*, controlling for the first 20 principal components of ancestry, was almost perfectly correlated ($r = .91$) with the percentage with which the SES indicators mediated the relation between European ancestry and *g*. That is, the more an SES indicator was genetically related to *g*, the more it mediated the relation between European ancestry and *g*. These results confirm that SES, genetics, and ancestry are confounded in this sample.

3.6.5 Brain volume and eduPGS

For subsequent analyses, we restricted focus to brain volume, because it is more plausibly a causal mediator of the relation between genetic ancestry and intelligence than is intracranial volume. In these analyses, we first examine if brain volume is predicted by our two eduPGS. We would expect this if variability in brain volume is related, genetically, to variability in *g* (within and between SIRE groups). Table 14 shows the results from the regression analysis for eduPGS and brain volume. Model 1a and 2a include sex, the cubic spline of age, and height, along with site- and family-fixed effects. We do not include SES, which includes parental education, since this is a criterion of eduPGS. Model 1b and 2b added the first twenty genetic PCs to control for population structure. As can be seen, both eduPGS metrics were statistically significantly related to brain volume, in the full sample, with and without controls for population structure.

Table 14. Regression results for the effects of polygenic scores on brain volume in the full sample without and with control for first 20 principal components (PCs) of genetic structure

	MTAG-eduPGS				Poly-eduPGS			
	Model 1a: no PCs		Model 1b: with PCs		Model 2a: no PCs		Model 2b: with PCs	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.26 (0.33)	0.436	0.33 (0.32)	0.302	0.18 (0.33)	0.588	0.31 (0.32)	0.344
Sex [M]	0.94 (0.02)	**<0.001**	0.94 (0.02)	**<0.001**	0.94 (0.02)	**<0.001**	0.94 (0.02)	**<0.001**
Age’	-0.08 (0.04)	**0.029**	-0.08 (0.03)	**0.015**	-0.07 (0.05)	0.051	-0.08 (0.03)	**0.019**
Age’’	0.00 (0.05)	0.960	-0.01 (0.05)	0.839	0.00 (0.05)	0.974	-0.01 (0.05)	0.814
Height	0.15 (0.01)	**<0.001**	0.16 (0.01)	**<0.001**	0.15 (0.01)	**<0.001**	0.16 (0.01)	**<0.001**
MTAG-eduPGS	0.27 (0.01)	**<0.001**	0.09 (0.01)	**<0.001**				
Poly-eduPGS					0.26 (0.01)	**<0.001**	0.09 (0.01)	**<0.001**
Random effects								
σ^2	0.28		0.27		0.28		0.27	
τ_{00}	$0.39_{\text{site\_id\_1:rel\_family\_id}}$ $0.01_{\text{site\_id\_1}}$		$0.36_{\text{site\_id\_1:rel\_family\_id}}$ $0.01_{\text{site\_id\_1}}$		$0.39_{\text{site\_id\_1:rel\_family\_id}}$ $0.01_{\text{site\_id\_1}}$		$0.36_{\text{site\_id\_1:rel\_family\_id}}$ $0.01_{\text{site\_id\_1}}$	
ICC	0.59		0.57		0.60		0.58	
N	$22_{\text{site\_id\_1}}$ $8600_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $8599_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $8600_{\text{rel\_family\_id}}$		$22_{\text{site\_id\_1}}$ $8599_{\text{rel\_family\_id}}$	
Observations	10245		10244		10245		10244	
Marginal R^2/ Conditional R^2	0.308/0.716		0.364/0.716		0.305/0.719		0.364/0.730	

Notes: Unstandardized beta coefficients (*b*) and *p*-values (*p*) from the mixed-effects models, with recruitment site and family common factors treated as random effects are shown. The values in parentheses are standard errors. The marginal and conditional R^2 of the mixed effects model are shown at the bottom. Model 1b and 2b additionally control for the first twenty genetic PCs. ICC = intraclass correlation coefficient. Note, only linear age was used in context to European ancestry for reasons discussed in the Methods section.

Table 15 summarizes the effects (betas) for the eduPGS on brain volume for the full sample and SIRE subsamples. The complete results are shown in Tables S7a to S7e of Supplementary File 1.

Table 15. Validities (*b*) from the mixed-effects regression models for eduPGS and brain volume in the full sample and in the SIRE subsamples

Predictor	Criterion	Controls	All	White	Hispanic	Black	Other
MTAG-eduPGS	Brain volume		**0.27**	**0.10**	**0.19**	**0.08**	**0.21**
MTAG-eduPGS	Brain volume	20PCs	**0.09**	**0.10**	**0.11**	0.01	**0.09**
Poly-eduPGS	Brain volume		**0.26**	**0.09**	**0.17**	**0.12**	**0.23**
Poly-eduPGS	Brain volume	20PCs	**0.09**	**0.09**	**0.11**	**0.08**	**0.14**
N			10245	5858	2005	1642	740

Note: Statistically significant results ($p < .05$) are presented in bold.

For Hispanics, Blacks, and the Other group, controlling for genetic ancestry components attenuates the associations somewhat. However, except in the case of MTAG-eduPGS in relation to Blacks, the effects remain statistically significant when controlling for genetic principal components, as would be expected were eduPGS causally related to brain volume.

3.6.6 Mediation by brain volume

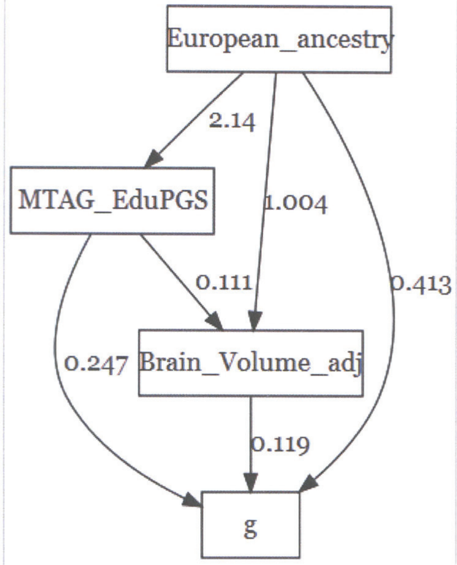
We next report the path model results. In the following analysis, brain volume is adjusted for the effects of sex, the cubic spline of age, height, and site- and family-related fixed effects. As seen in Figure 6a/b, both eduPGS and brain volume partially explain the association between European ancestry and *g*. Moreover, the path between eduPGS and *g* is partially mediated by brain volume. Note that in these models the effects are calculated by multiplying the coefficients for each path and summing the products. For example, the effects of European ancestry on *g* by way of brain volume is calculated as $1.004 * 0.119$ (for European ancestry * brain volume * *g*) + $2.14 * 0.111 * 0.119$ (for European Ancestry * MTAG-eduPGS * Brain volume * *g*) = 0.15. The full model fit results along with the results for each SIRE subsample are provided in Tables S9a-S9j of Supplemental File 1.

It is theoretically possible that our eduPGS are only capturing global genetic European ancestry effects. To test this conjecture, we ran a robustness analysis using pseudo-eduPGS, created by weighing random SNPS with the eduPGS beta weights. If eduPGSs are only acting as proxies for global genetic ancestry, these pseudo eduPGS should produce similar results because in this scenario, it would not matter if the SNPs were causally associated with the trait. The results of this analysis are shown in Supplementary File 1, Tab 17. For the ten individual pseudo-PGS, path associations between European ancestry and pseudo-PGS varied around zero (ranging from $b = -1.13$ to 1.05). The average of the ten showed no ancestry-related association. Neither the average nor the individual pseudo-PGS predicted *g* or brain volume. So, our eduPGS are not merely acting as proxies for global genetic ancestry. Of course, magnitudes of eduPGS differences may still be biased between groups for other reasons, as pointed out by others (e.g., Yair & Coop, 2022).

A number of researchers (e.g., Nisbett et al., 2012; Warne, 2021) have objected to the ancestry-brain volume argument using an analogy with sex differences. According to this argument, there are large sex differences in brain volume favoring males, yet little to no general intelligence ones. It is reasoned that, therefore, ancestry-associated differences in brain volume do not need to be related to cognitive differences. Because sex differences in general intelligence are trivial in the ABCD sample, we could directly evaluate this hypothesis by repeating the path model (Figure 6b), but with two alternations. We changed European ancestry to dummy-coded sex (Female = 1, Male = 0). Moreover, we replaced *g* scores, which were adjusted for sex in the MGCFA model, with age-corrected NIHTBX composite scores ($N = 9{,}904$ in this subsample), which were unadjusted for sex.

Girls had trivially higher ($d = .02$) NIHTBX scores overall. Figure 7 shows the path model, with the full results provided in the Supplementary File. The path model for sex differences shows that brain volume has a suppression effect on the female cognitive advantage. Conditioned on brain volume, girls score higher than boys, but the average brain size for girls was smaller and, as such, girls score about the same as boys. So, brain volume is indeed related to sex differences in general intelligence. Whereas brain volume has a suppression effect on sex differences, it is a mediator of the cognitive advantage for the group with European ancestry. This is because African and Amerindian ancestry is associated with lower, not higher, intelligence independent of brain volume. Full model results are provided in Supplementary File, Figure S16.

a. Path model using MTAG-eduPGS


European_ancestry
2.14
MTAG_EduPGS
1.004
0.111
0.413
0.247
Brain_Volume_adj
0.119
g

b. Path model using poly-eduPGS

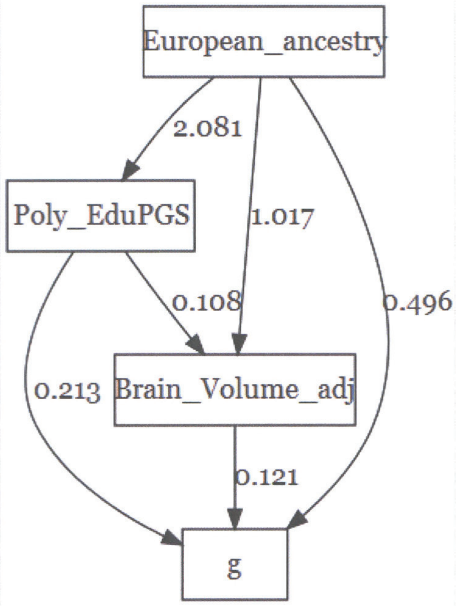

European_ancestry
2.081
Poly_EduPGS
1.017
0.108
0.496
0.213
Brain_Volume_adj
0.121
g

Figure 6. Path diagram for the relation between European ancestry, MTAG-eduPGS, adjusted brain volume, and *g* in the full sample. $N = 10{,}245$. Full model results are provided in the Supplementary File

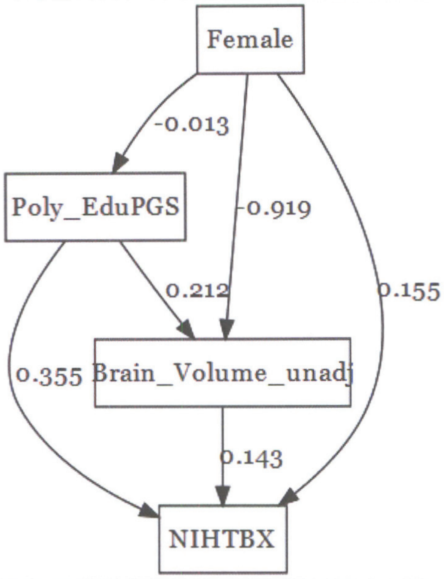


Figure 7. Path diagram for the relation between sex, Poly-eduPGS, adjusted brain volume and *g* in the full sample, $N = 9{,}904$

The results for the mediation analyses are summarized in Table 16. The full results are provided in Tables S10a-S10t of Supplementary File 1. All proportions were statistically significant for all samples. Notably, brain volume mediated 11-23% of the European ancestry and *g* association and 6-13% of the association between poly-eduPGS and *g*. Poly-eduPGS, in turn, explained 14-25% of the relation between European ancestry and brain volume.

Jensen (1998, *p*. 442) argued that brain size could explain 40% of the Black-White SIRE differences based on an assumed within-population correlation, between brain volume and *g*, of $r =$ .40. In contrast, we find that brain volume can explain only approximately 15% of the association between European genetic ancestry and *g*. This is because the beta for brain volume with respect to *g* is relatively low within our sample of 10-year-old children. It is notable, though, that the portion of ancestry difference explained corresponds approximately to the within-SIRE effect of brain volume on *g*. This suggests that if the brain-volume-by-*g* correlation increases with age, then a proportionately larger amount of the ancestry-by-*g* relationship could be explained by brain volume. Because the ABCD follows the children for ten years it will be possible to test this conjecture in the future.

Table 16. Summary of the mediation results for the full sample and SIRE subsamples

			Proportion of association mediated (%)				
Predictor	**Mediator/ component**	**Criterion**	**All**	**White**	**Hispanic**	**Black**	**Other**
European ancestry	Brain volume	*g*	**15.50**	**18.62**	**11.36**	**17.49**	**23.20**
European ancestry	Poly-eduPGS	*g*	**37.40**	**48.64**	**22.80**	**25.10**	**88.60**
European ancestry	Poly-eduPGS	Brain volume	**17.10**	**15.76**	**14.18**	**14.18**	**25.01**
Poly-eduPGS	Brain volume	*g*	**8.23**	**5.68**	**9.38**	**12.50**	**8.52**
N			10245	5858	2005	1642	740

Note: Statistically significant results ($p < .05$) are presented in bold.

3.7 Method of correlated vectors

Table 17 reports the vectors of *g* loadings, eduPGS effects, European ancestry effects, brain volume effects, heritabilities, and standardized Black-White (B-W) and Hispanic-White (H-W) group differences for each of the cognitive tests.

Table 17. Vectors for MCV analysis

Subtests	Subtest reliability	*g*-loading	Poly eduPGS	MTAG eduPGS	European ancestry	Brain volume	h^2	B-W gap	H-W gap
Picture vocabulary	0.800	0.612	0.206	0.228	1.351	0.121	0.338	0.940	0.520
Flanker	0.830	0.386	0.045	0.070	0.534	0.053	0.238	0.360	0.120
List sorting	0.770	0.559	0.162	0.169	0.858	0.092	0.384	0.630	0.270
Card sorting	0.810	0.433	0.075	0.098	0.601	0.046	0.335	0.490	0.210
Pattern comparison	0.730	0.338	0.045	0.058	0.308	0.015	0.205	0.240	0.080
Picture sequence memory	0.840	0.267	0.086	0.126	0.458	0.024	0.434	0.530	0.150
Oral reading recognition	0.900	0.600	0.189	0.223	0.957	0.145	0.497	0.660	0.300
Matrix test	0.850	0.510	0.164	0.188	0.828	0.099	0.175	0.590	0.300
Little man test		0.228	0.063	0.088	0.194	0.070	0.345	0.270	0.050
RVLT immediate	0.760	0.530	0.120	0.151	0.634	0.023	0.461	0.530	0.150
RVLT delay	0.660	0.540	0.121	0.137	0.588	0.018	0.259	0.530	0.170

Notes: h^2 = cognitive test heritability estimates; B-W gap = standardized Black-White differences; H-W gap = standardized Hispanic-White differences.

Table 18 shows the Pearson correlations between the vectors. (Table S12c in the Supplementary File 1, additionally shows the Spearman correlations, but these were more or less the same.) The results uncorrected for cognitive test reliabilities appear above the diagonal; those corrected for test reliabilities appear below. Note the latter are based on one less test, owing to missing data. As can be seen, all correlations are moderately to strongly positive, with the exception of the correlation between *g* loadings and heritability ($r = .17$ to $.21$) and heritability and brain volume ($r = .08$ to $.21$). For comparison, the literature reports a bare-bones meta-analytic correlation of $r = .31$ ($SD_r = .43$) ($k = 12$) between *g* loadings and heritability (te Nijenhuis et al., 2019). A possible explanation for these low correlations with heritabilities is that this sample consists of 10-year-old children and at young ages the specific heritability estimates of cognitive tests is relatively high compared to the heritability conditioned on *g* (Procopio et al., 2022). Alternatively, it is likely that the statistical imprecision in estimation of ACE biases the correlations towards zero. Supplementary File 1, Table S18 reports published vector correlations; the correlations in Table 18 generally agree with those. Moreover, the results are in line with previous findings that correlations with both European genetic ancestry and eduPGS are highly correlated with *g* loadings.

The factor analysis of the eight vectors produced a single factor, which explained no less than 71 percent of the variance. The factor loadings are provided in the final column of Table 18. EduPGS, genetic ancestry, and the SIRE group differences had the highest loadings on this common factor.

Table 18. Matrices of correlations between vector correlations, uncorrected and corrected for test reliability along with loadings on the common factor for the vectors

	Vector correlations (*r*s)								Factor loadings
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	
[1] *g* loading	1.00	0.84	0.79	0.86	0.45	0.17	0.78	0.76	0.86
[2] Poly-eduPGS	0.82	1.00	0.99	0.90	0.65	0.32	0.90	0.88	0.97
[3] MTAG-eduPGS	0.77	0.99	1.00	0.88	0.63	0.40	0.91	0.86	0.96
[4] European ancestry	0.76	0.90	0.88	1.00	0.63	0.20	0.95	0.97	0.97
[5] Brain volume	0.42	0.63	0.59	0.68	1.00	0.08	0.47	0.65	0.68
[6] Heritability	0.21	0.32	0.39	0.21	0.21	1.00	0.33	0.12	0.31
[7] B-W gap	0.71	0.90	0.90	0.94	0.46	0.36	1.00	0.95	0.95
[8] H-W gap	0.64	0.85	0.83	0.97	0.69	0.11	0.92	1.00	0.95

Note: Correlations of cognitive test scores with *g* loadings, the effects of poly-eduPGS, MTAG-eduPGS, % European ancestry, and adjusted brain volume on subtest scores, and the Black-White (B-W) and Hispanic-White (H-W) gaps. *N* = 11 subtests above the diagonal (with corrections for subtest reliability), and *N* = 10 below diagonal (with no corrections for subtest reliability).

Additionally, Table 19 shows the results, controlling alternatively for the vector of *g* loadings and the vector of eduPGS effects using partial correlations. As seen, when *g* loadings are controlled for, the positive manifold mostly persists. This implies that the positive manifold is not solely due to a common association with *g* loadings. In contrast, when eduPGS effects are controlled for, the vector correlations often turn null or negative. In other words, it is the effect of genes, and not of phenotypic *g*, which is the glue which holds together the manifold of positive correlations.

Finally, the average within-SIRE effects of MTAG-eduPGS and PolyeduPGS on tests for Blacks and Whites, Hispanics and Whites, and Others and Whites correlated very strongly with, respectively, the Black-White, Hispanic-White, Other-White standardized mean differences ($r = .81$ to $r = .91$). After controlling for *g* loadings, the correlations were still high to very high ($r = .57$ to $r = .78$). These results are reported in Tables S12d and S12e of Supplementary File 1.

Table 19. Results from method of correlated vectors analysis controlling for Poly-eduPGS (above the diagonal) and controlling for *g* loadings (below the diagonal)

	[2]	[3]	[4]	[5]	[6]	[7]	[2]
[1] *g* loadings	.42	-.22	-.18	.08	.09	-.40	.42
[2] European ancestry	1	.16	-.21	.74	.89	-.09	1
[3] Brain volume adj.	.54	1	-.17	-.33	.23	-.12	.54
[4] Subtest heritability	.11	.01	1	.12	-.33	.55	.11
[5] B-W difference	.88	.23	.32	1	.76	.21	.88
[6] H-W difference	.97	.54	-.01	.87	1	-.09	.97
[7] MTAG-eduPGS	.65	.49	.44	.76	.65	1	.65
[8] Poly-eduPGS	.65	.56	.32	.74	.68	.97	.65

3.8. Group-level analyses

The descriptive statistics for the group-level analyses are provided in Supplementary File 2. Table 20 shows the descriptive statistics and Table 21 shows the correlation matrices. As can be seen, the correlations between the various PGS and between PGS and NIHTBX were very high across both SIRE/national groups and religious groups. This is because we are dealing with ecological correlations,

Table 20. Descriptive statistics for religious & SIRE/national group-level analyses

Group	Sample size	NIH TBX	Lee GWAS Lead	Lee MTAG eduPGS	Lee MTAG cogPGS	Lee MTAG mathclass	Lee MTAG mathability	Hill MTAG IQPGGS	Weissbrod Poly eduPGS
	N	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*
Agnostic	238	0.50	0.47	0.47	0.45	0.45	0.41	0.26	0.47
Atheist	249	0.69	0.60	0.70	0.68	0.68	0.64	0.24	0.62
Buddhist	36	0.19	0.15	0.23	0.25	0.25	0.26	0.41	0.09
Evangelical Protestant	813	-0.29	-0.38	-0.35	-0.32	-0.33	-0.28	-0.07	-0.32
Greek Orthodox Christian	71	-0.05	0.09	0.13	0.10	0.13	0.13	-0.13	0.06
Hindu	34	0.25	0.21	0.28	0.16	0.27	0.23	0.10	0.45
Historically Black Church	190	-0.68	-1.31	-1.31	-1.19	-1.21	-1.04	-0.37	-1.23
Jehovahs Witness	69	-0.55	-0.62	-0.62	-0.58	-0.54	-0.45	-0.17	-0.64
Jewish	223	0.66	0.73	0.81	0.71	0.72	0.58	0.33	0.65
Mainline Protestant	1268	0.27	0.34	0.34	0.34	0.32	0.29	0.15	0.32
Mormon	890	0.23	0.47	0.50	0.49	0.48	0.44	0.12	0.51
Muslim	75	-0.21	-0.18	-0.13	-0.13	-0.14	-0.12	0.12	-0.32
Not identified	539	-0.43	-0.53	-0.53	-0.49	-0.49	-0.42	-0.19	-0.50
Nothing in particular	2452	0.09	0.06	0.04	0.03	0.04	0.03	0.00	0.04
Other Christian	1952	-0.27	-0.42	-0.44	-0.40	-0.41	-0.36	-0.14	-0.41
Roman Catholic	2098	-0.04	0.12	0.11	0.08	0.10	0.07	0.01	0.10
Something else	177	-0.11	-0.21	-0.28	-0.25	-0.25	-0.21	-0.07	-0.21
Unitarian	100	0.55	0.42	0.49	0.46	0.46	0.41	-0.03	0.49
Asian Indian	53	0.35	0.25	0.23	0.11	0.21	0.17	0.20	0.48
Black & other Puerto Rican	90	-0.52	-0.73	-0.75	-0.74	-0.75	-0.72	-0.32	-0.62
Black African	59	-0.27	-1.27	-1.22	-1.08	-1.11	-0.92	-0.31	-1.30
Black Caribbean	51	-0.31	-1.35	-1.35	-1.17	-1.20	-0.96	-0.31	-1.25
Central & South American	352	-0.06	-0.14	-0.14	-0.20	-0.15	-0.17	0.02	-0.13
Chinese	81	0.90	0.62	0.75	0.57	0.77	0.67	0.49	0.15
Dominican	38	-0.26	-0.41	-0.39	-0.38	-0.39	-0.37	-0.01	-0.32
Filipino	51	0.43	0.19	0.06	-0.13	0.02	-0.10	0.30	-0.28
Korean & Japanese	33	0.65	0.39	0.43	0.35	0.49	0.50	0.52	0.13
N. Africa & Mid. East	47	0.13	0.33	0.42	0.38	0.40	0.37	0.18	0.20
Native American	39	-0.41	-0.36	-0.37	-0.36	-0.32	-0.26	-0.03	-0.37
NH Black & White	418	-0.20	-0.58	-0.55	-0.50	-0.50	-0.42	-0.18	-0.54
Not Identified	217	-0.15	-0.11	-0.15	-0.14	-0.13	-0.10	-0.11	-0.17
Other Asian	52	0.41	0.40	0.44	0.34	0.41	0.31	0.32	0.34
Other Cuban	30	-0.20	-0.43	-0.30	-0.27	-0.23	-0.16	-0.21	-0.53
Other Hispanic	518	-0.21	-0.46	-0.43	-0.42	-0.40	-0.36	-0.11	-0.44
Other Mexican	460	-0.44	-0.44	-0.43	-0.50	-0.43	-0.45	-0.16	-0.39
Pacific Islander	17	0.02	0.10	0.10	-0.03	0.13	0.12	-0.15	0.00
USA Black	1499	-0.78	-1.41	-1.37	-1.22	-1.25	-1.04	-0.38	-1.33
Vietnamese	24	0.24	0.28	0.38	0.28	0.41	0.37	0.63	-0.30
White	5858	0.28	0.49	0.47	0.45	0.44	0.39	0.13	0.48
White & Asian Indian	44	0.83	0.58	0.63	0.57	0.61	0.55	0.19	0.57
White & Chinese	77	0.68	0.67	0.62	0.50	0.61	0.54	0.23	0.46
White & Filipino	60	0.61	0.58	0.56	0.46	0.54	0.48	0.24	0.13
White & Korean/Japanese	78	0.65	0.65	0.67	0.63	0.68	0.63	0.64	0.42
White & Native American	144	0.00	0.23	0.18	0.18	0.17	0.15	-0.02	0.25
White & Pacific Islander	25	0.28	0.46	0.36	0.21	0.28	0.14	0.09	0.27
White Cuban	151	-0.20	0.15	0.07	0.03	0.02	-0.03	0.03	0.07
White Mexican	775	-0.23	-0.15	-0.16	-0.22	-0.17	-0.20	-0.09	-0.14
White Puerto Rican	133	-0.17	0.03	-0.04	-0.08	-0.08	-0.13	-0.09	-0.18

as a result the individual-level variability is removed (Robinson, 1950). However, the average correlations between edu/IQ-PGS and cognitive ability was lower across SIRE/national groups (r = .89) than across religious ones (r = .94). So, edu/IQ-PGS do less well at predicting cognitive ability across SIRE/national groups. This is consistent with the hypothesis of linkage-disequilibrium-related confounding in the context of ancestral groups. This is because American SIRE/national groups are more related to ancestry than American religious groups are.

Table 21. Correlation between average polygenic scores across SIRE and national groups (N = 30) and across religious groups (N = 18)

SIRE and national groups								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[1] NIHTBX	1.00	0.89	0.90	0.90	0.91	0.92	0.90	0.83
[2] Lee GWASlead-eduPGS		1.00	1.00	0.99	0.99	0.98	0.86	0.97
[3] MTAG-eduPGS			1.00	1.00	1.00	0.99	0.87	0.97
[4] Lee MTAG-cognitivePGS				1.00	1.00	0.99	0.86	0.97
[5] Lee MTAG-mathclassPGS					1.00	1.00	0.88	0.96
[6] Lee MTAG-mathabilityPGS						1.00	0.89	0.95
[7] Hill MTAG-IQPGS							1.00	0.77
[8] Poly-eduPGS								1.00
Religious groups								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[1] NIHTBX	1.00	0.95	0.95	0.95	0.95	0.95	0.88	0.95
[2] Lee GWASlead-eduPGS		1.00	1.00	1.00	1.00	0.99	0.89	0.99
[3] MTAG-eduPGS			1.00	1.00	1.00	1.00	0.90	0.99
[4] Lee MTAG-cognitivePGS				1.00	1.00	1.00	0.90	0.99
[5] Lee MTAG-mathclassPGS					1.00	1.00	0.90	0.99
[6] Lee MTAG-mathabilityPGS						1.00	0.90	0.99
[7] Hill MTAG-IQPGS							1.00	0.87
[8] Poly-eduPGS								1.00

Next we look at the regression plots for eduPGS and NIHTBX scores. Those for MTAG-eduPGS and Poly-eduPGS are shown in Figure 8, while the plots for the other five eduPGS are provided in Supplementary File 2, Figures S1-S7. As seen, religious groups are mostly within the 95% confidence intervals (in gray) of the regression lines for both PGS. The situation is somewhat different for SIRE/national groups. Notably, Black African and Caribbean immigrant children fall conspicuously below the regression lines for both MTAG and Poly-eduPGS. This same pattern is also found for math PGS, though less so for IQPGS. Second, East Asians also fall below the regression lines especially for Poly-eduPGS, though this pattern is found less so for math PGS and the opposite is found for IQpgs. Generally, the results are suggestive of some ancestry related bias. Nonetheless, the correlations between SIRE/national group-level edu/math/IQ PGS and cognitive ability are quite strong. Unlike examples discussed by Martin et al. (2017), we find a substantial degree of consistency between PGS and phenotype in context to SIRE/national groups (with a variance explained ranging from R^2 = .68 to .85).

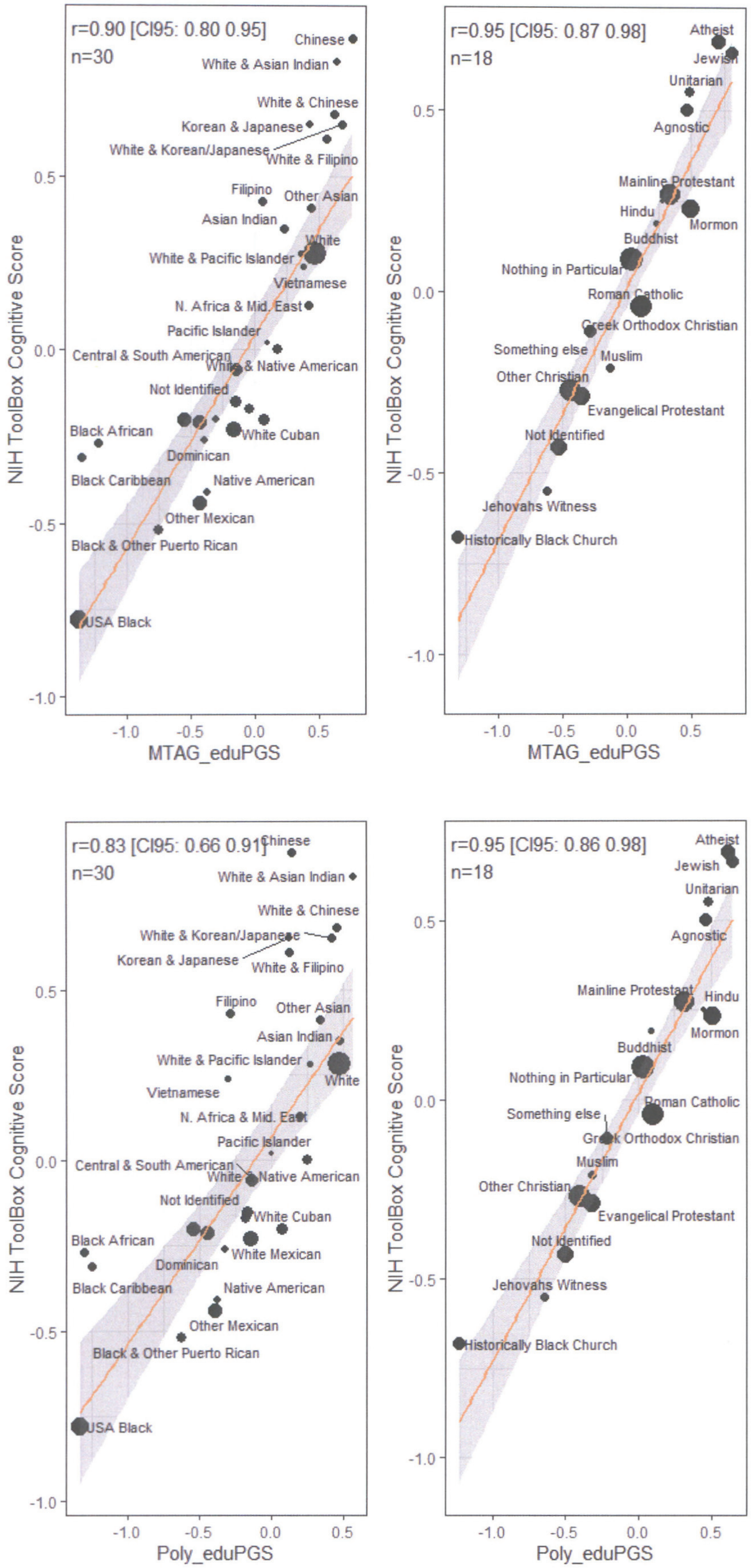


Figure 8. Regression plots for NIH Toolbox Composite Scores by MTAG-eduPGS or by Poly-eduPGS by SIRE (left) and Religions (right) groups

3.9. MRI-based brain predictor, cognitive ability, and ancestry

The full results for the MRI-based brain predictor analyses are provided in Supplementary File 2. Table 22 reports the validities based on the mixed-effects regression models. As can be seen, the MRI-based brain predictor is positively correlated with our *g* scores ($r = .38$ to $r = .48$) within all SIRE groups and, moreover, European ancestry is positively associated with the MRI-based brain predictor in all SIRE groups. In these samples, the mean B/W, H/W, and O/W *d*-values were, respectively, $d = .90$, .45, and .51.

Table 23 summarizes the results for the mediation analyses. The full results are provided in Supplementary File 2. All proportions were statistically significant for all samples. Notably, the MRI-based brain predictor mediated 28.8 to 48.9% of the European ancestry and *g* association, while MTAG eduPGS statistically mediated 35.57 to 57.97% of the relation between European ancestry and the brain predictor. So, using a broad MRI-based predictor of intelligence, instead of whole brain volume, increases, not decreases the portion of variance explained by MRI data (38.50% vs. 15.50%).

Table 22. Validities (*b*) from the mixed-effects regression models for the MRI-brain predictor, g, and genetic ancestries in the full sample and in the SIRE subsamples. Controlled variables are sex, age, height, and 20 genetic principal components

	Criterion	All	White	Hispanic	Black	Other
Brain predictor	*g*	**0.41**	**0.38**	**0.42**	**0.46**	**0.48**
European ancestry	Brain predictor	**0.98**	**0.64**	**0.79**	**0.86**	**0.75**
African ancestry	Brain predictor	**-0.96**	-0.52	**-0.69**	**-0.71**	**-0.78**
Amerindian ancestry	Brain predictor	**-1.14**	-0.83	**-0.84**	-0.28	-0.76
N		7464	4484	1462	999	519

Notes: Statistically significant results ($p < .05$) are presented in bold.

Table 23. Summary of the mediation results for the full sample and SIRE subsamples using the MRI-based brain predictor variable

			Proportion of association mediated				
Predictor	Mediator/component	Criterion	All	White	Hispanic	Black	Other
European ancestry	Brain predictor	*g*	**38.50**	**28.79**	**29.30**	**43.30**	**48.90**
European ancestry	MTAG-eduPGS	Brain predictor	**37.50**	**57.97**	**40.10**	**46.23**	**35.57**
N			7464	4484	1462	999	519

4. Discussion

4.1 General discussion

In the US, there are medium-to-large general mental ability differences between Whites, Hispanics, and Blacks. Warne (2021) recently argued for a model of partial genetic differences. We closely examined Warne's (2021) arguments and the papers he cited. Based on this careful reading, we formulated 15 hypotheses.

4.1.1 Measurement of group differences and Spearman's hypothesis

Hypothesis 1 stated there would be medium-to-large mean differences between American SIRE groups in *g*, eduPGS, and brain/intracranial volume. The data showed that this was strongly the case. The only exception was that Hispanics in our sample showed a small-to-medium sized difference in *g* ($d = .38$), which is smaller than typically reported. Overall, the group differences in the ABCD sample were strongly comparable to what was found in previous studies.

Moreover, our MGCFA analysis confirmed Spearman's hypothesis, that the cognitive differences between SIRE groups are mainly due to differences in *g*. This finding increases the plausibility of a genetic hypothesis because within SIRE groups, *g* differences are more heritable and neurology-associated than non-*g* cognitive differences. Moreover, many commonly proposed sources of SIRE differences, such as differences in schooling, do not increase *g*, but instead increase domain-specific cognitive abilities (e.g., Ritchie & Bates, 2015).

4.1.2 Measurement invariance for groups with European genetic ancestry

Hypothesis 2 predicted that strict measurement invariance (SFI) would hold across the full range of European genetic ancestry, and this was found to be the case. This outcome is not entirely surprising because strict measurement invariance already held across the four SIRE groups, which differ markedly in European genetic ancestry.

The finding of SFI is relevant to the debate about the cause of group differences because SFI implies that the sources of between-group differences are a subset of the sources of within-group differences and not influences which are more or less unique to one or another group (Lubke et al., 2003). Specific examples of previously proposed sources of SIRE group differences in cognitive ability which have been found to violate SFI, as they are due to group unique causes, include those causing the Flynn effect (Wicherts et al., 2004) and those causing stereotype threat (Wicherts et al., 2005). Because the sources of the Flynn effect and stereotype threat violate SFI, they are not plausible causal candidates for American SIRE differences, which do not violate SFI. So, the set of possible environmental explanations of SIRE and ancestry differences in *g* is constrained by the finding of SFI (Warne, 2021).

4.1.3 Within-group heritability and shared environmentality

Hypothesis 3 concerned the heritability and shared environmentality of intelligence and brain/intracranial volume in the full sample and within SIRE groups. In this national sample of ten-year-old children, the heritability estimates of general intelligence were moderate-to-high in both the White and the combined non-Whites sample, while the shared environmentality estimates were low. This is in line with the findings of Pesta et al. (2020, 2023), who report that heritabilities are similar across White and non-White groups. Moreover, both brain and intracranial volume were highly heritable in the full sample, the White sample, and the non-White sample. The only other comparable SIRE-specific volumetric estimates we are aware of are those reported by Rushton and Osborne (1995) for cranial capacity. Those authors found White and Black heritabilities of $h^2 = .56$ (111 pairs) and $h^2 = .31$ (125 pairs), respectively, for their age-, height-, and sex-adjusted cranial capacities. The heritabilities in the present sample are higher.

Hypothesis 4 was whether the heritability and shared environmentality estimates for the White and combined non-White groups were similar in magnitude. For brain and intracranial volume, the heritability estimates were somewhat higher, while the shared environmentalities were a bit lower for the non-White group. For NIHTBX and *g*, the heritability estimates were somewhat higher, while the shared environmentalities were lower for the White group. Generally, these results suggest that adverse environments are not substantially depressing the heritabilities in the non-White sample as predicted by some (e.g., Scarr-Salapatek, 1971; van den Oord & Rowe, 1997).

While some have argued that within-group heritability is not necessarily relevant to between group differences (e.g., Greenspan, 2022), within-group environmentality, or 1 minus heritability, is relevant insofar as group differences are due to a subset of the factors that cause differences within groups (Sesardic, 2005). If these factors are weakly related to differences in *g* within groups, then differences need to be large between groups. For example, we find that shared environment correlates at $r = .35$ with *g* in the combined White-Hispanic-Black-Other sample. So, to explain the 1.35 *d* advantage associated with 100% European ancestry in contrast to Amerindian and African ancestry, one would need $d = 1.35 / .35 = 3.86$ *SD* of shared environmental effect, a huge effect. Note, in these calculations we do not take into account the correlation between non-shared environment and *g*. This is because non-shared environmental factors seem to be unlikely candidates for the association between ancestry and *g* at the population level because, after all, ancestry and the factors it proxies are mostly shared between biological siblings.

4.1.4 Admixture regression for *g*

Hypothesis 5 concerned if genetic ancestry explained SIRE differences in *g*; this was found to be the case, in line with recent research results (e.g., Fuerst, 2021; Kirkegaard et al., 2019; Lasker et al., 2019; Warne, 2020). In the full sample, SIRE variables had no validity independent of genetic ancestry, despite that "SIRE acts as a surrogate to an array of social, cultural, behavioral, and environmental variables" (Fang et al., 2019). Moreover, within SIRE groups, European genetic ancestry was positively related to *g*, while Amerindian and African ancestries were negatively so. These results imply that broad socio-cultural factors which have similar effects on members of the same SIRE group regardless of ancestry (e.g., SIRE-based segregation and discrimination, cultural norms, dialect, language, involuntary minority status) do not cause the intelligence gaps between SIRE groups.

A general factor of SES statistically explained approximately 40% of the ancestry effect found in the full sample. However, SES is both genetically and environmentally related to child intelligence within groups. Indeed, consistent with the findings of Krapohl and Plomin (2016), the correlation between child eduPGS and parental SES was about the same as that between child eduPGS and child cognitive ability (Supplemental File 2, Table S10). Thus, SES could be controlling for either genetic or shared environmental effects between groups.

Two previous studies using the ABCD sample found that various variables did not account for the association between cognitive ability and European ancestry, namely state-level racism, child-reported experiences of discrimination, parental ethnic attachment, and child skin, hair and eye color, with the latter capturing effects of possible color-based discrimination (Fuerst, 2021; Fuerst et al., 2021). However, there are potentially other factors (e.g., premature birth, perinatal drug use, and lead exposure) which could potentially statistically explain ancestry-associated differences. As a result, "it is important to interpret associations between ancestry and … outcome with caution, because unmeasured environmental confounders may still explain the effect." (González Burchard et al., 2005). While a number of researchers have argued that genetic and environmental influences are hopelessly confounded in global admixture analysis (e.g., Conley & Fletcher, 2017), our argument is that you can use path analysis to investigate plausible pathways for future research using more powerful methods.

4.1.5 Polygenic scores and *g*

Hypothesis 6 was whether eduPGS predicted intelligence within SIRE groups and within sibships. This was found to be the case. Consistent with previous results (Howe et al., 2022), we found a within-sibship shrinkage, relative to the association within the full sample, of between 20-27%. Generally, this finding supports the conclusion that eduPGS is causally related to *g* within SIRE groups.

4.1.6 Brain volume, *g*, ancestry, and eduPGS

Hypothesis 7 was if brain volume differences are related to *g* and to eduPGS within SIRE groups. Within SIRE groups and within sibships, brain volume was related to *g*.

Hypothesis 8 was whether brain/intracranial volume differences between SIRE groups are explained by ancestry-associated genetic variation. We found that SIRE differences were completely accounted for by genetic ancestry. African and Amerindian ancestry both had a large negative effect on brain volume, while European ancestry had a positive effect. For intracranial volume, the effects were moderate but the same patterns were found.

Hypothesis 9 was if brain/intracranial volume correlated with genetic ancestry within SIRE groups. This was found to be the case. Generally, the results, at least with respect to African ancestry, agree with those from older studies, which report that either brain size/volume or cranial capacity is related to ancestry among admixed African American populations (e.g., Bean, 1906; Herskovits, 1969; Pearl, 1934).

Hypothesis 10 was whether SES would be related to *g* and brain volume among biological children but not adoptive children subsamples. It was found that SES was uncorrelated with both *g* and brain volume among adopted children, while it was positively correlated with both variables among biological children. While these results are not particularly strong, owing to the small size and a lack of pre-adoption information (e.g., information about contact adoptees have with biological parents, prenatal and pre-adoptive conditions, etc.), they are consistent with previously reported ones which show that parental SES is only weakly related to adoptive child IQ (e.g., McGue et al., 2007; Beaver et al., 2014). The current findings tentatively suggest that the relation between parental SES and *g,* in this sample, is mainly genetic in origin. Owing to low power, and thus relatively large standard errors, a clear conclusion was not possible for parental SES and the brain volume of adopted children.

Hypothesis 11 was if brain volume differences were predicted by eduPGS in all SIRE groups. It was found that eduPGS predicted brain volume in 19/20 of the models, and this included 9/10 of the models with controls for the first 20 principal components. The sole exception involved MTAG-eduPGS, but not the more causally relevant poly-eduPGS, in relation to African Americans when controls for the first PCs were also included. The interpretation in this one case is uncertain, because controlling for genetic PCs may correct for causal genetic effects (Lawson et al., 2020). Considering the overall consistency of the results, this one statistically non-significant finding could be a fluke due to insufficient power. However, future research including other MRI-based variables could examine if eduPGS robustly predict neurological features among African Americans. Generally, the results found confirm that eduPGS are related to brain volume and also that brain volume is related to *g* in all SIRE groups.

Hypotheses 12 and 13 concerned whether the relation between European genetic ancestry and *g* was mediated both by eduPGS and brain volume and if the pathway running from European ancestry through eduPGS to *g* was also mediated by brain volume. The association between European genetic ancestry and *g* was partially statistically accounted for by brain volume. Moreover, the pathways between European ancestry, polygenic scores and *g* were also mediated by brain volume. This later finding supports the interpretation that ancestry-related eduPGS differences are causally related to ancestry-associated differences in *g*.

Various authors argue that there is a lack of relationship between brain size differences and sex differences in intelligence and that ancestry differences in brain size might likewise be unrelated to intelligence differences (Nisbett et al., 2012; Warne, 2021). However, we find that, between sexes, brain size is indeed related to intelligence; specifically, the smaller brain volume of girls moderates the female cognitive advantage. As there is no corresponding African or Amerindian ancestry advantage, brain volume acts, instead, as a mediator with respect to ancestry.

4.1.7 Method of correlated vectors

Hypothesis 14 was if a positive manifold of effects across tests would be found for *g*-loadings, SIRE/genetic ancestry, eduPGS, and brain volume. MCV provided corroborating support for a model according to which ancestry and SIRE group differences are partially genetic. The vectors of cognitive test effects associated with genetic, brain, and ancestral/SIRE variables, exhibited a positive manifold as would be predicted if the variables captured common *g*-related genetic effects. Moreover, the magnitude of ancestry-adjusted effects of eduPGS within SIRE groups correlated with SIRE group differences, even when controlling for *g* loadings.

To see if this latter finding replicated, we re-analyzed data from Fuerst et al. (2021; Table S5). Again, we find that eduPGS-loadings within SIRE groups correlate with both the vectors of ancestry effects and SIRE differences independent of *g* loadings (*r*s = .77 to .87). This phenomenon, which is predicted by a genetic hypothesis, is in need of an explanation by a non-genetic hypothesis. These results are of additional interest because it is known that many environmental effects do not produce positive correlations with genetic variables and *g*-loadings (te Nijenhuis et al., 2019, Table 1; Warne, 2021).

That said, MCV is a crude method and the power to test the strength of correlations is low (Kan et al., 2013). This issue could be investigated further using biometric decomposition of phenotypic mean differences (Jensen, 1998, *pp*. 464-467) or variants of this method. Another solution is to carry out several replications, and to carry out a series of meta-analyses and correct for well-known measurement errors (Jensen, 1998).

Finally, Hypothesis 15 was whether there would be a higher-order putative genetic factor resulting from the intercorrelating genetic variables and SIRE/genetic ancestry in our study. This was found to be the case, though our heritability estimates loaded weakly on this factor. This was perhaps due to the unreliability inherent in these estimates, which had wide confidence intervals.

4.1.8 Additional analyses

We found that education, math, and cognitive PGS correlated strongly with one another across SIRE and national groups. Moreover, these PGSs correlated strongly with NIH Toolbox cognitive scores. So, the results do not suggest a large discordance between PGS and measured cognitive ability. Nonetheless, the correlations for SIRE and national groups were lower than those for religious groups, suggesting some bias. Moreover, specific groups like East Asians and immigrant Black

Africans generally had higher measured cognitive scores than predicted by PGS. This finding is consistent with the presence of some degree of ancestry-related confounding, although the overall correlations are also consistent with the hypothesis that PGS differences correspond, imperfectly, with SIRE and national genetic differences causal of phenotypic differences. So, more research is needed to determine to what extent observed PGS differences are causative of general intelligence differences (Fuerst et al., 2021).

We also found that a neurological-based predictor of cognitive ability, created using machine learning, explained a substantially larger portion of the association between European ancestry and *g* than did whole brain volume. Moreover, this predictor had a higher association with eduPGS than did brain volume. So, while it is true that, as Warne (2021, *p*. 492) notes, "evidence is mounting that brain organization and functioning may be more important biological determinants of intelligence than sheer brain size," it does not appear that taking into account more aspects of brain organization and functioning diminishes the association between ancestry and *g*.

4.2 Summary and limitations of the study

We found that cognitive differences between SIRE groups were primarily in *g*. Moreover, strict factorial invariance held both between SIRE groups and along the full range of European ancestry. These general cognitive ability differences were moderately to highly heritable within both the White and the combined non-White SIRE groups. The *g* differences between SIRE groups were completely explained by genetic ancestry. Further, the relation between genetic ancestry and *g* was substantially explained by two variables: brain volume, and polygenic scores for education. Moreover, the magnitude of SIRE and ancestry-associated differences strongly correlated with the effects of eduPGS within SIRE groups. These findings are consistent with a genetic hypothesis.

A number of findings presented here have been fairly well replicated. These include: 1) mean differences and Spearman's hypothesis, 2) measurement invariance between SIRE groups, 3) high within-group heritability and low within-group shared environmentality, 4) the association between European genetic ancestry and cognitive ability, 5) the zero to low correlation between SES and cognitive ability among adoptees, and 6) within-sibship validity of eduPGS and brain volume. However, other findings are not well-replicated. First, only two other studies, using the same sample, examined the relation between the within-SIRE effect of eduPGS on tests and the magnitude of cognitive differences between them (Fuerst et al., 2021; Lasker et al., 2019).

Second, while there are large independent bodies of research on polygenic scores and brain volume, brain volume and intelligence, and ancestry-related differences in brain volume, the present study is the only one which has analyzed all these variables at once using mediation and path analysis. In our view, Warne (2021) was generally justified to find the brain volume data uncompelling at the time because no one had previously shown, using a larger representative sample, that 1) there are MRI-measured brain volume differences between SIRE groups, 2) brain volume has high heritability in White and non-White groups, 3) SIRE differences in brain volume are completely explained by genetic ancestry, 4) eduPGS predicts brain volume within SIRE groups, and 5) brain volume mediates the relation between genetic ancestry and cognitive ability.

Third, while some studies have looked at the relations between eduPGS and national average cognitive scores (e.g., Piffer, 2015, 2019), only one other study has examined the association between eduPGS and SIRE scores within a country using cognitive and genetic data from the same sample (Fuerst, Hu, & Connor, 2022). So, replication of these latter results is needed before we can be confident that there is something in need of explaining.

Regarding eduPGS, a number of researchers have called into question the trans-ancestral-group portability of these (e.g., Harden, 2021). We do find that eduPGS are portable in the sense of being predictive in non-White SIRE groups. However, because our eduPGS are based on European-descent individuals, it is uncertain to what extent eduPGS differences correspond to causally relevant genetic differences between SIRE groups and ancestries. We find, despite some results consistent with some bias against African and East Asian ancestry, that the educational and cognitive PGS differences generally predict SIRE/national differences.

This finding is consistent with a scenario according to which eduPGS differences between ancestries partially reflect causally relevant genetic ones which cause cognitive differences. However, one could also imagine that eduPGS differences are indexing causally irrelevant genetic variation, due to confounding related to population structure and also that these eduPGS differences are coincidentally correlated with intelligence differences for various historical reasons (Fuerst, Kirkegaard & Piffer, 2021). While this hypothesis is less parsimonious than a causal hypothesis, it has some plausibility. Researchers who believe that eduPGS differences across ancestries do not correspond at all with causally relevant differences should try to formulate specific, testable hypotheses for the causes of these observed differences and their covariance with phenotypic and neurological differences.

Regarding brain volume, this study builds upon previous admixture-regression results by incorporating neurological data. We use whole-brain volume as the neurological predictor due to its evolutionary significance and also because of the large amount of research relating this variable to both intelligence and evolutionary history. However, whole-brain volume is a crude index of intelligence-related neurological features (Warne, 2021). This line of research could be extended by developing a better neurological predictor of intelligence, applying machine learning to the rich set of structural and functional MRI variables in the ABCD dataset; there is a substantial body of relevant research (e.g., Pohl et al., 2019). In supplemental analyses, we ran models using one multimodal MRI-based brain predictor, based on over 50,000 neurological variables, and found that we could account for a substantially larger percentage of the association between European ancestry and cognitive ability. Future research can attempt to develop better MRI-based predictors of intelligence to determine to what extent the relation between ancestry and *g* is explainable neurologically.

Regarding admixture regression, this study focuses on the relation between global or overall admixture and intelligence. As recognized in the genetic epidemiology literature, global admixture analysis is a weaker method than local admixture analysis. With local admixture analysis, one can examine specific regions of admixed genomes that are associated with a trait. Global admixture analyses are preliminary first steps.

4.3 Related research

Fuerst & Kirkegaard (2016) find that European (versus African and Amerindian) global genetic ancestry strongly predicts regional socioeconomic and cognitive differences across the Americas. This finding suggests that the individual-level association between genetic ancestry and cognitive ability may not be localized only in the US. Additionally, Bird (2021) finds that, between African and European populations, differentiation in SNPs related to both education and intelligence is commensurate with average SNP differentiation. The author takes the findings as evidence against a model of strong divergent selection for intelligence. Of relevance, the magnitude of reported differentiation (Fst = .111 to .126) would imply moderate to large genetically-caused cognitive differences given the heritabilities found in the present sample (computations shown in Appendix C). Finally, Norris et al. (2019) find signatures of ancestry-based assortative mating on educational SNPS in admixed American populations. As the authors note, this implies that genetic variants causally related to education differ between ancestries. So, presently known differentiation in SNPs related to education and intelligence may, indeed, track to some degree causally-relevant genetic variation between ancestries.

4.4 Future research

Those who find a partial genetic hypothesis plausible can follow the next logical step, namely carrying out local-admixture analyses to examine the regions of the genome where the association between intelligence and ancestry is most pronounced (Kirkegaard et al., 2019; Lasker et al., 2019). Local-admixture analysis can differentiate between three competing hypotheses. These are: 1) a shared-environment account in which case cognitive ability is only passively correlated with genetic ancestry; in this case, local admixture should not have more explanatory power than global admixture; 2) a simple phenotype-based-discrimination account in which case cognitive ability is reactively correlated with genetic ancestry by way of local regions associated with stereotypical racial aspects of phenotypes (such as skin color); and 3) an alternative account, in which case cognitive ability is correlated with *g* by way of diffuse regions of the local genome. Support for 1) or 2) would strengthen a non-genetic model. If it is found that local admixture is associated with *g*, and if this association is not isolated to regions of the genome which code for racially stereotypic aspects of phenotypes, then a follow-up study could examine if the association between local ancestry and intelligence is pronounced in genomic regions associated with neurological functioning as predicted by a genetic model.

4.5 Conclusion

Taken together, these different lines of evidence provide convergent genetically- and neurologically-informed evidence for a partial genetic hypothesis. To paraphrase Sandra Scarr (1981, *p.* 528), holes can be poked in each line of evidence, but when the evidentiary lines are laid on top of one another, like sheets of paper, the holes do not go all the way through. Researchers, of course, may disagree. Many did with Scarr's view about individual differences in IQ, made in the early 1980s, long before the genomic revolution (Scarr, 1981). Those who disagree with our interpretation should try to account for the results taken jointly, not piecemeal, and identify environmental variables which mimic the results found.

Author contributions: All individual-level analyses were conducted by JGRF under the supervision of B. J. Pesta while at Cleveland State University (2020-2021). VS and EOWK helped write and edit the paper and check the tables.

Acknowledgments: Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data, but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the research results and interpretations of the authors alone and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Version 3.01. The raw data are available at https://nda.nih.gov/edit_collection.html?id=2573. Additional support for this work was made possible from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680), and Children and Screens: Institute of Digital Media and Child Development Inc. NDA waived the requirement of creating an NDA data project (NDA, Oct 26, 2022).

References

Akshoomoff, N., Newman, E., Thompson, W.K., McCabe, C., Bloss, C.S., Chang, L., ... & Jernigan, T.L. (2014). The NIH Toolbox Cognition Battery: Results from a large normative developmental sample (PING). *Neuropsychology* 28(1): 1-10.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.

Bates, T.C., Neale, M.C. & Maes, H.H. (2019). umx: A library for structural equation and twin modelling in R. *Twin Research and Human Genetics* 22: 27-41.

Beals, K.L, Smith, C.L, Dodd, S.M, Angel, J.L, Armstrong, E., Blumenberg, B., Girgis, F.G., Turkel, S., Gibson, K.R., … & Trinkaus, E. (1984). Brain size, cranial morphology, climate, and time machines. *Current Anthropology* 25: 301-330.

Bean, R.B. (1906). Some racial peculiarities of the Negro brain. *American Journal of Anatomy* 5: 353-432.

Beaver, K.M., Schwartz, J.A., Al-Ghamdi, M.S., Kobeisy, A.N., Dunkel, C.S. & van der Linden, D. (2014). A closer look at the role of parenting-related influences on verbal intelligence over the life course: Results from an adoption-based research design. *Intelligence* 46: 179-187.

Betjemann, R.S., Johnson, E.P., Barnard, H., Boada, R., Filley, C.M., Filipek, P.A., Willcutt, E.G., DeFries, J.C. & Pennington, B.F. (2010). Genetic covariation between brain volumes and IQ, reading performance, and processing speed. *Behavior Genetics* 40: 135-145.

Bird, K.A. (2021). No support for the hereditarian hypothesis of the Black–White achievement gap using polygenic scores and tests for divergent selection. *American Journal of Physical Anthropology* 175: 465-476.

Bollen, K.A. & Pearl, J. (2013). Eight myths about causality and structural equation models. In: S.L. Morgan (ed.), *Handbook of Causal Analysis for Social Research*, pp. 301-328. Springer.

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics* 96: 37-53.

Chande, A. T., Nagar, S. D., Rishishwar, L., Mariño-Ramírez, L., Medina-Rivas, M. A., Valderrama-Aguirre, A. E., ... & Gallo, J. E. (2021). The impact of ethnicity and genetic ancestry on disease prevalence and risk in Colombia. *Frontiers in Genetics*, *12*, 1624.

Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J. (2014). *Multivariate analysis of mixed data: The R Package PCAmixdata*. arXiv.

Chen, Y., Ide, J.S., Li, C.S., Chaudhary, S., Le, T.M., Wang, W., Zhornitsky, S., Zhang, S. & Li, C.R. (2022). Gray matter volumetric correlates of dimensional impulsivity traits in children: Sex differences and heritability. *Human Brain Mapping* 43: 2634-2652.

Cheng, C.Y., Reich, D., Haiman, C.A., Tandon, A., Patterson, N., Elizabeth, S., ... & Kao, W.L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: a genetic admixture analysis in three US population cohorts. *PLoS One* 7(3): e32840.

Choi, S.W., Mak, T.S.H. & O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols* 15: 2759-2772.

Choquet, H., Yin, J. & Jorgenson, E. (2021). Cigarette smoking behaviors and the importance of ethnicity and genetic ancestry. *Translational Psychiatry* 11: 120.

Clark, G. & Henneberg, M. (2021). Interpopulational variation in human brain size: Implications for hominin cognitive phylogeny. *Anthropological Review* 84: 405-429.

Cochran, G. & Harpending, H. (2009). *The 10,000 Year Explosion: How Civilization Accelerated Human Evolution*. Basic Books.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Erlbaum.

Conley, D. & Fletcher, J. (2017). *The Genome Factor*. Princeton Univ. Press.

Connor, G. & Fuerst, J.G.R. (2022). Linear and partially linear models of behavioral trait variation using admixture regression. In: R. Lynn (ed.), *Intelligence, Race and Sex: Some Controversial Issues. A Tribute to Helmuth Nyborg at 85*. Budapest, Hungary: Arktos.

Coon, C.S. (1982). *Racial Adaptations*. Nelson-Hall.

Decker, S.L. (2021). Don't use a bifactor model unless you believe the true structure is bifactor. *Journal of Psychoeducational Assessment* 39: 39-49.

DeFries, J.C. (1972a). Quantitative aspects of genetics and environment in the determination of behavior. In: L. Ehrman, G. Omenn & E. Caspari (eds.), *Genetics, Environment, and Behavior*, pp. 6-16. San Diego, CA: Academic Press.

Dolan, C.V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research* 35: 21-50.

Elliott, M.L., Belsky, D.W., Anderson, K., Corcoran, D.L., Ge, T., Knodt, A., Prinz, J.A., Sugden, K., … & Hariri, A.R. (2019). A polygenic score for higher educational attainment is associated with larger brains. *Cerebral Cortex* 29: 3496-3504.

Fan, C.C., Bartsch, H., Schork, A.J., Chen, C.H., Wang, Y., Lo, M.T., Brown T.T., Kuperman, J.M., … & Dale, A.M. (2015). Modeling the 3D geometry of the cortical surface with genetic ancestry. *Current Biology* 25: 1988-1992.

Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., Vujkovic, M., Scott, M.D., … & Tang, H. (2019). Harmonizing genetic ancestry and race/ethnicity in genome-wide association studies. *American Journal of Human Genetics* 105: 763-772.

Fernández, J.R. & Shiver, M.D. (2004). Using genetic admixture to study the biology of obesity traits and to map genes in admixed populations. *Nutrition Reviews* 62(S2): S69-S74.

Flynn, J.R. (2018). Academic freedom and race: You ought not to believe what you think may be true. *Journal of Criminal Justice* 59: 127-131.

Flynn, J.R., te Nijenhuis, J. & Metzen, D. (2014). The g beyond Spearman's *g:* Flynn's paradoxes resolved using four exploratory meta-analyses. *Intelligence* 44: 1-10.

Fox, J. & Weisberg, S. (2018). *An R Companion to Applied Regression*, 2nd ed. Sage.

Frisby, C.L. & Beaujean, A.A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence* 51: 79-97.

Fuerst, J.G. (2021). Robustness analysis of African genetic ancestry in admixture regression models of cognitive test scores. *Mankind Quarterly* 62: 396-413.

Fuerst, J. & Kirkegaard, E.O. (2016). Admixture in the Americas: Regional and national differences. *Mankind Quarterly* 56: 255-373.

Fuerst, J.G., Hu, M. & Connor, G. (2021). Genetic ancestry and general cognitive ability in a sample of American youths. *Mankind Quarterly* 62: 186-216.

Fuerst, J.G.R., Kirkegaard, E.O. & Piffer, D. (2021). More research needed: There is a robust causal vs. confounding problem for intelligence-associated polygenic scores in context to admixed American populations. *Mankind Quarterly* 62: 151-185.

González-Burchard, E., Borrell, L.N., Choudhry, S., Naqvi, M., Tsai, H.J., Rodriguez-Santana, J.R., ... & Risch, N. (2005). Latino populations: A unique opportunity for the study of race, genetics, and social environment in epidemiological research. *American Journal of Public Health* 95: 2161-2168.

Gottfredson. L.S. (2007). Flynn, Ceci, and Turkheimer on race and intelligence: Opening moves. *Cato Unbound,* November 26.

Greenspan, N.S. (2022). Genes, heritability, 'race', and intelligence: Misapprehensions and implications. *Genes* 13(2): 346.

Guo, G. & Wang, J. (2002). The mixed or multilevel model for behavior genetic analysis. *Behavior Genetics* 32: 37-49.

Halder, I., Matthews, K.A., Buysse, D.J., Strollo, P.J., Causer, V., Reis, S.E. & Hall, M.H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep* 38: 1185-1193.

Hambly, W.D. (1947). Cranial capacities, a study in methods. *Fieldiana. Anthropology* 36(3): 25-75.

Harden, K.P. (2021). *The Genetic Lottery: Why DNA Matters for Social Equality*. Princeton Univ. Press.

Hatton, S. (2018). Preview of the ABCD Study® Data Release 3.0. Department of Neurosciences, UC San Diego. https://abcdstudy.org/wp-content/uploads/2020/06/ABCD_Release3.0.pdf

Heeringa, S.G. & Berglund, P.A. (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. bioRxiv.

Herskovits, M.J. (1969). *The Anthropometry of the American Negro*, vol. 11. Ardent Media.

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C. & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research* 51: 257-258.

Hill, W.D., Marioni, R.E., Maghzian, O., Ritchie, S.J., Hagenaars, S.P., McIntosh, A.M., ... & Deary, I.J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry* 24: 169-181.

Hood, S.B. (2010). *Latent variable realism in psychometrics*. Doctoral dissertation, Indiana University. https://scholarworks.iu.edu/dspace/handle/2022/8120

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., ... & Davies, N. M. (2021). Withinsibship GWAS improve estimates of direct genetic effects. BioRxiv. https://doi.org/10.1101/2021.03.05.433935

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., ... & Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, *54*(5), 581-592.

Hu, M., Lasker, J., Kirkegaard, E.O. & Fuerst, J.G. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych* 1(1): 240-261.

Iacono, W.G., Heath, A.C., Hewitt, J.K., Neale, M.C., Banich, M.T., Luciana, M.M., Madden, P.A., Barch, D.M. & Bjork, J.M. (2018). The utility of twins in developmental cognitive neuroscience research: How twins strengthen the ABCD research design. *Developmental Cognitive Neuroscience* 32: 30-42.

Imai, K., Keele, L., Tingley, D. & Yamamoto, T. (2014). Comment on Pearl: Practical implications of theoretical results for causal mediation analysis. *Psychological Methods* 19: 482-487.

Jansen, A.G., Mous, S.E., White, T., Posthuma, D. & Polderman, T.J. (2015). What twin studies tell us about the heritability of brain development, morphology, and function: A review. *Neuropsychology Review* 25: 27-46.

Jensen, A.R. (1998). *The g Factor: The Science of Mental Ability*. Praeger.

Judd, N., Sauce, B., Wiedenhoeft, J., Tromp, J., Chaarani, B., Schliep, A., van Noort, B., Penttilä, J., … & Klingberg, T. (2020). Cognitive and brain development is independently influenced by socioeconomic status and polygenic scores for educational attainment. *Proceedings of the National Academy of Sciences* 117: 12411-12418.

Kan, K.J. (2012). *The nature of nurture: the role of gene-environment interplay in the development of intelligence*. Doctoral dissertation, Universiteit van Amsterdam. UvA-DARE (Digital Academic Repository).

Kan, K.J., Wicherts, J.M., Dolan, C.V. & van der Maas, H.L. (2013). On the nature and nurture of intelligence and specific cognitive abilities: The more heritable, the more culture dependent. *Psychological Science* 24: 2420-2428.

Kirkegaard, E.O.W., & Fuerst, J.G.R. (2023). A multimodal MRI-based predictor of intelligence and its relation to socio-demographic characteristics. *Mankind Quarterly* 63: 374-397.

Kirkegaard, E.O., Wang, M. & Fuerst, J.G. (2017). Biogeographic ancestry and socioeconomic outcomes in the Americas: A meta-analysis. *Mankind Quarterly* 57: 398-427.

Kirkegaard, E.O.W., Woodley, M.A., Williams, R.L., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych* 1(1): 1-25.

Krapohl, E. & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry* 21: 437-443.

Lasker, J. (July, 2019). A surfeit of evidence on Spearman's hypothesis. *20th Annual Meeting of the International Society for Intelligence Research*, Minneapolis, Minnesota.

Lasker, J., Pesta, B.J., Fuerst, J.G. & Kirkegaard, E.O. (2019). Global ancestry and cognitive ability. *Psych* 1(1): 431-459.

Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., ... & Timpson, N.J. (2020). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* 139: 23-41.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, T., … & Cesarini, D. (2018). Gene discovery and polygenic prediction from a 1.1-million person GWAS of educational attainment. *Nature Genetics* 50: 1112–1121.

Lee, J.J., McGue, M., Iacono, W.G., Michael, A.M. & Chabris, C.F. (2019). The causal influence of brain size on human intelligence: Evidence from within-family phenotypic associations and GWAS modeling. *Intelligence* 75: 48-58.

Lewis, J.E., DeGusta, D., Meyer, M.R., Monge, J.M., Mann, A.E. & Holloway, R.L. (2011). The mismeasure of science: Stephen Jay Gould versus Samuel George Morton on skulls and bias. *PLoS Biology* 9(6): e1001071.

Long, J.A. (2020). *jtools: Analysis and presentation of social scientific data* (R package version 2.1.0) [Data set]. https://cran.r-project.org/package=jtools.

Lubke, G.H., Dolan, C.V., Kelderman, H. & Mellenbergh, G.J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence* 31: 543-566.

Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., ... & Kenny, E.E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *American Journal of Human Genetics* 100: 635-649.

McClearn, G.E. & DeFries, J.C. (1973). *Introduction to Behavioral Genetics*. San Francisco: Freeman.

McGue, M., Keyes, M., Sharma, A., Elkins, I., Legrand, L., Johnson, W. & Iacono, W.G. (2007). The environments of adopted and non-adopted youth: Evidence on range restriction from the Sibling Interaction and Behavior Study (SIBS). *Behavior Genetics* 37: 449-462.

Mehanna, M., McDonough, C.W., Smith, S.M., Gong, Y., Gums, J.G., Chapman, A.B., ... & Cooper-DeHoff, R.M. (2022). Influence of genetic West African ancestry on metabolomics among hypertensive patients. *Metabolites* 12: 783.

Mehta, C.M., Malins, J.G., Noble, K.G. & Gruen, J.R. (2017). Cortical area and subcortical volume mediate the effect of parental education and adverse experiences on cognitive performance in youth. bioRxiv.

Murray, A.L. & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence* 41: 407-422.

Murray, C. (2021). *Facing Reality: Two Truths about Race in America*. Encounter Books.

Nisbett, R.E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D.F. & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist* 67: 130-159.

Norris, E.T., Rishishwar, L., Wang, L., Conley, A.B., Chande, A.T., Dabrowski, A.M., ... & Jordan, I.K. (2019). Assortative mating on ancestry-variant traits in admixed Latin American populations. *Frontiers in Genetics* 10: 359

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., ... & Esko, T. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54(4), 437-449.

Pearl, R. (1934). The weight of the Negro brain. *Science* 80: 431-434.

Pesta, B.J., Kirkegaard, E.O., te Nijenhuis, J., Lasker, J. & Fuerst, J.G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence* 78: 101408.

Pesta, B.J., Fuerst, J.F. & Shibaev, V. (2021). Research practices in differential psychology: Comments and a correction to Winston. *Азимут научных исследований: педагогика и психология* 10(2): 395-398.

Pesta, B.J., te Nijenhuis, J., Lasker, J., Kirkegaard, E.O. & Fuerst, J.G. (2023). On group differences in the heritability of intelligence: A reply to Giangrande and Turkheimer (2022). *Intelligence* 98: 101737.

Pietschnig, J., Gerdesmann, D., Zeiler, M. & Voracek, M. (2022). Of differing methods, disputed estimates, and discordant interpretations: The meta-analytical multiverse of brain volume and IQ associations. *Royal Society Open Science* 9(5): 211621.

Piffer, D. (2015). A review of intelligence GWAS hits: Their relation to country IQ and the issue of spatial autocorrelation. *Intelligence* 53: 43-50.

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from GWAS hits: A replication of previous findings using recent data. *Psych* 1(1): 55-75.

Plomin, R. & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics* 19: 148-159.

Pohl, K.M., Thompson, W.K., Adeli, E. & Linguraru, M.G. (eds.) (2019, October 13). *Adolescent brain cognitive development neurocognitive prediction. Proceedings of the first challenge, ABCD-NP 2019*, held in conjunction with MICCAI. Shenzhen, China.

Procopio, F., Zhou, Q., Wang, Z., Gidziela, A., Rimfeld, K., Malanchini, M. & Plomin, R. (2022). The genetics of specific cognitive abilities. *Intelligence* 95: 101689.

Relethford, J.H. (2010). Population-specific deviations of global human craniometric variation from a neutral model. *American Journal of Physical Anthropology* 142: 105-111.

Rindermann, H. (2018). *Cognitive Capitalism: Human Capital and the Wellbeing of Nations*. Cambridge Univ. Press.

Rindermann, H., Becker, D. & Coyle, T.R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence* 78: 101406.

Ritchie, S.J., Bates, T.C. & Deary, I.J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology* 51: 573-582.

Robinson, W.S. (1950) Ecological correlations and the behaviour of individuals. *American Sociological Review* 15: 351-357.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(2): 1-36.

Roth, P.L., van Iddekinge, C.H., DeOrtentiis, P.S., Hackney, K.J., Zhang, L. & Buster, M.A. (2017). Hispanic and Asian performance on selection procedures: A narrative and meta-analytic review of 12 common predictors. *Journal of Applied Psychology* 102: 1178-1202.

Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression-unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences*, *26(*2), 381-390.

Rushton, J.P. & Ankney, C.D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience* 119: 692-732.

Rushton, J.P. & Jensen, A.R. (2005). "Thirty years of research on race differences in cognitive ability": Correction to Rushton and Jensen (2005). *Psychology, Public Policy, and Law* 11: 406.

Rushton, J.P. & Osborne, R.T. (1995). Genetic and environmental contributions to cranial capacity in Black and White adolescents. *Intelligence* 20: 1-13.

Rushton, J.P. & Rushton, E.W. (2003). Brain size, IQ, and racial-group differences: Evidence from musculoskeletal traits. *Intelligence* 31: 139-155.

Scarr, S. (1981). Having the last word. In: S. Scarr (ed.), *Race, Social Class, and Individual Differences in IQ*, pp. 523-532. Hillsdale NJ: Erlbaum.

Scarr-Salapatek, S. (1971). Race, social class, and IQ: Population differences in heritability of IQ scores were found for racial and social class groups. *Science* 174: 1285-1295.

Scheiber, C. (2016a). Do the Kaufman tests of cognitive ability and academic achievement display construct bias across a representative sample of Black, Hispanic, and Caucasian school-age children in grades 1 through 12? *Psychological Assessment* 28: 942-952.

Scheiber, C. (2016b). Is the Cattell–Horn–Carroll-based factor structure of the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) construct invariant for a representative sample of African-American, Hispanic, and Caucasian male and female students ages 6 to 16 years? *Journal of Pediatric Neuropsychology* 2(3): 79-88.

Sesardic, N. (2005). *Making Sense of Heritability*. Cambridge Univ. Press.

Smith, C.L. & Beals, K.L. (1990). Cultural correlates with cranial capacity. *American Anthropologist* 92: 193-200.

Sun, H., Lin, M., Russell, E.M., Minster, R.L., Chan, T.F., Dinh, B.L., ... & Chiang, C.W. (2021). The impact of global and local Polynesian genetic ancestry on complex traits in Native Hawaiians. *PLoS Genetics* 17: e1009273.

te Nijenhuis, J., Choi, Y.Y., van den Hoek, M., Valueva, E. & Lee, K.H. (2019). Spearman's hypothesis tested comparing Korean young adults with various other groups of young adults on the items of the Advanced Progressive Matrices. *Journal of Biosocial Science* 51: 875-912.

Thompson, W.K., Barch, D.M., Bjork, J.M., Gonzalez, R., Nagel, B.J., Nixon, S.J. & Luciana, M. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Developmental Cognitive Neuroscience* 36: 100606.

Tingley, D, Yamamoto, T, Hirose, K, Keele, L & Imai, K. (2013). Mediation: R package for causal mediation analysis (R package version 4.4.2) [Data set]. http://CRAN.R-project.org/package=mediation.

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., ... & Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence* 42: 83-88.

Van den Oord, E.J. & Rowe, D.C. (1997). An examination of genotype-environment interactions for academic achievement in an US national longitudinal survey. *Intelligence* 25: 205-228.

Van Leeuwen, M., Peper, J.S., van den Berg, S.M., Brouwer, R.M., Pol, H.E.H., Kahn, R.S. & Boomsma, D.I. (2009). A genetic analysis of brain volumes and IQ in children. *Intelligence* 37: 181-191.

Vuoksimaa, E., Panizzon, M.S., Chen, C.H., Fiecas, M., Eyler, L.T., Fennema-Notestine, C., … & Kremen, W.S. (2015). The genetic association between neocortical volume and general cognitive ability is driven by global surface area rather than thickness. *Cerebral Cortex* 25: 2127-2137.

Warne, R.T. (2020a). Continental genetic ancestry source correlates with global cognitive ability score. *Mankind Quarterly* 60: 400-422.

Warne, R.T. (2020b). *In the Know: Debunking 35 Myths about Human Intelligence*. Cambridge Univ. Press.

Warne, R.T. (2021). Between-group mean differences in intelligence in the United States are >0% genetically caused: Five converging lines of evidence. *American Journal of Psychology* 134: 480-501.

Weintraub, S., Dikmen, S.S., Heaton, R.K., Tulsky, D.S., Zelazo, P.D., Slotkin, J., ... & Gershon, R. (2014). The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: Validation in an adult sample. *Journal of the International Neuropsychological Society* 20: 567-578.

Weiss, L.G. & Saklofske, D.H. (2020). Mediators of IQ test score differences across racial and ethnic groups: The case for environmental and social justice. *Personality and Individual Differences* 161: 109962.

Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., ... & Price, A.L. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nature Genetics* 52: 1355-1363.

Weissbrod, O., Kanai, M., Shi, H., Gazal, S., Peyrot, W.J., Khera, A.V.,... & Price, A.L. (2022). Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nature Genetics* 54: 450-458.

Wicherts, J.M., Dolan, C.V., Hessen, D.J., Oosterveld, P., Van Baal, G.C.M., Boomsma, D.I. & Span, M.M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence* 32: 509-537.

Wicherts, J.M., Dolan, C.V. & Hessen, D.J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology* 89: 696-716.

Wiers, C.E., Towb, P.C., Hodgkinson, C.A., Shen, P.H., Freeman, C., Miller, G., ... & Volkow, N.D. (2018). Association of genetic ancestry with striatal dopamine D2/D3 receptor availability. *Molecular Psychiatry* 23: 1711-1716.

Yair, S. & Coop, G. (2022). Population differentiation of polygenic score predictions under stabilizing selection. *Philosophical Transactions of the Royal Society* B 377: 20200416.

Zaidi, A.A. & Mathieson, I. (2020). Demographic history mediates the effect of stratification on polygenic scores. *Elife* 9: e61548.

CHAPTER 13

A MULTIMODAL MRI-BASED PREDICTOR OF INTELLIGENCE AND ITS RELATION TO RACE/ETHNICITY

EMIL O.W. KIRKEGAARD AND JOHN G.R. FUERST

Abstract

We used data from the Adolescent Brain Cognitive Development Study to create a multimodal MRI-based predictor of intelligence. We applied the elastic net algorithm to over 50,000 neurological variables. We find that race can confound models when a multiracial training sample is used, because models learn to predict race and use race to predict intelligence. When the model is trained on non-Hispanic Whites only, the MRI-based predictor has an out-of-sample model accuracy of $r = .51$, which is 3 to 4 times greater than the validity of whole brain volume in this dataset. This validity generalized across the major socially-defined racial/ethnic groupings (White, Black, and Hispanic). There are race gaps on the predicted scores, even though the model is trained on White subjects only. This predictor explains about 37% of the relation between both the Black and Hispanic classification and intelligence.

Keywords: machine learning, elastic net, race, intelligence, MRI

1. Introduction

Whole brain volume, as measured by Magnetic Resonance Imaging (MRI), and intelligence are known to correlate (Haier, 2016; Pietschnig et al., 2022). This is as expected based on what is known about the evolution of human brain size and intelligence (DeSilva et al., 2021). However, it is also known that whole brain volume is only a small part of the explanation for human differences in intelligence. For instance, a large study of British adults using Structural Equation Modeling (SEM) found a latent correlation of about $r = .28$ between whole brain volume and intelligence (Cox et al., 2019). As this result is based on SEM, it avoids random measurement error, a major confounding factor in previous analyses of the relations between brain size and intelligence (Gignac & Bates, 2017). Taking this value as the current best estimate and using the usual path squaring rule, this equates to about 10% of the variance in intelligence explained by whole brain volume alone.

Multi-modal and fine-grained analyses are needed to better explain intelligence differences (Kievit et al., 2012). We searched for the best published multivariable model for intelligence. One of the best predictive models so far published was able to achieve a correlation of $r = .67$ using functional MRI data in the Human Connectome Project dataset (Sripada et al., 2018). However, this study has still not been formally published after four years. There are two additional studies claiming very high accuracies. First, Wang et al. (2015) used a small sample of children ($N = 164$) and found a cross-validated correlation of $r = .72$. Given overfitting issues with small samples, and the inability of other research teams to achieve such results in larger samples, it may not replicate. Second, a team of Chinese researchers reported cross-validated correlations of $r = .45$ and $r = .51$ across multiple smaller samples in two studies (Jiang et al., 2020a,b). In contrast to these results, teams who took part in the 2019 ABCD Neurocognitive Prediction Challenge to build a structural-MRI predictor of fluid intelligence achieved much weaker results: r's = .15, .14, .28, .21, .11, and .19. (Brueggeman et al., 2019; Li et al., 2019; Ren et al., 2019; Valverde et al., 2019; Wlaszczyk et al., 2019; Zhang-James et al., 2019). These results were based on a proper hold-out sample that the researchers did not have access to while designing and training their models. As such, it seems like overfitting is a serious concern in this literature. Clearly much more research is needed.

In addition to intelligence, brain volume is also known to correlate with racial or ancestry group membership (Rushton & Ankney, 2009). However, it is not known to what extent race is related to multimodal MRI-predictors of intelligence. Additionally, it is not known if multimodal MRI-predictors of intelligence are equally predictive across racial groups. As it is, some research reports that the validity of specific MRI variables differs across socially-identified races (e.g., Zahodne et al., 2015). So, a multimodal MRI-predictor trained in one group may not work as well in another.

Furthermore, we wanted to move beyond whole brain volume towards a machine learning (ML) algorithm trained for optimal predictive validity of intelligence. So, in this paper, we apply machine learning to a rich set of MRI variables to create MRI-based predictors of intelligence. We then examine the relationship between the MRI-based predictors, intelligence, and socially-identified race. Based on past research, we make the following predictions:

1. Predictors based on all available MRI modalities will achieve validities for intelligence that are significantly higher than that for brain volumes alone.
2. MRI-based predictors will show race differences, and these MRI-based differences will statistically mediate the association between race and intelligence.
3. Validities for multimodal predictors, as for brain volume (Farias et al., 2012; Jensen & Johnson, 1994), will generalize across races, underscoring the neuropsychological unity of mankind.

2. Data & Methods

2.1. Data source

We used the 3.01 dataset of the ABCD study (Adolescent Brain Cognitive Development; https://nda.nih.gov/abcd). This is a publicly funded nationally representative sample of nearly 12,000 American children who were first interviewed at age 9-10. It contains high-quality microarray genomic data, multi-modal MRI data, psychological data, as well as a wide range of surveys (Casey et al., 2018; Hagler et al., 2019; Jernigan et al., 2018). Details about sampling are provided by Garavan et al. (2018).

The sample is 52.2% male and 47.8% female. We used a set of binary questions to determine socially-identified race/ethnicity (from now on: "social race"). To be clear, we call this variable *social race* to distinguish it from morphologically or genetically-identified race or ancestry (e.g., Fuerst, 2015; Sesardic, 2010).

In this case, each child's primary guardian, most often the mother, was asked whether the child belonged to a series of racial groups ("What race do you consider the child to be? Please check all that apply."): White/European American, Black/African American, American Indian, Alaska Native, Native Hawaiian, Guamanian, Samoan, Other Pacific Islander, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian, and Other. The primary guardian was also asked if the child was of Hispanic ethnicity (i.e., Spanish cultural origin). These questions are based on the 1997 Office of Management and Budget guidelines, according to which race may be thought of in terms of ancestry in addition to culture and explicitly refers to the geographic origin of one's ancestors (Office of Management and Budget, 1997).

While one can code the data in many ways, we were interested in relatively large groupings for our models. As such, we coded the data so we have categories for non-Hispanic Whites and non-Hispanic Blacks, an inclusive category for Hispanics (with any racial combination), and an Other category, which contains everybody else, including individuals who were identified as belonging to multiple racial categories. Table 1 shows the sample composition in terms of social race.

Table 1. Sample composition and mean intelligence by social race. *SD* = standard deviation. The total sample mean IQ was 95.4 with an *SD* of 16.9

Group	**Count**	**Percent**	**IQ Mean ± *SD***
White	6179	52.0	100.0 ± 15.0
Hispanic	2411	20.3	90.6 ± 15.9
Black	1782	15.0	81.2 ± 16.4
Other	1454	12.2	97.3 ± 17.0

2.2. Neurological data

The neurological data were pre-computed and released for use by the ABCD team (Casey et al., 2018). These data include 81 MRI files, containing a total of 52,785 MRI variables. Nineteen of these variables had too much (>20%) missing data and so were removed, leaving 52,766. These variables were often highly correlated (e.g., right and left hemisphere cortical volume), so the effective dimensionality is far less than suggested by the total number.

We excluded quality control variables and non-brain related variables (e.g., "Number of frames after removing dummy frames") from the neurological files, and focused only on variables directly related to brain structure, neural responses, or functional connectivity. As there were some outliers in the data, we winsorised the data to three median absolute deviations (i.e., robust standard deviations). To avoid sex and age confounding, we residualized all variables for age, sex, and height and the interaction between sex and age.[1] We explored a few different models for residualization, but we found that different residualization models had only minor effects on the results. This is not surprising given the small variation in age in this dataset ($M = 10.7$, $SD = 1.0$), and the relative lack of sex differences in body size and development at age 10 (e.g., boys were only 0.9 cm shorter on average).

Most of the functional MRI (fMRI) variables had about 17% missing data as those subjects did not attend those measurements. Because we wanted to use every available modality of the MRI data to maximize predictive validity, we excluded subjects with more than 5% missing data instead of, alternatively, dropping the fMRI variables. The remaining amount of missing data was near-zero and we imputed these data using median imputation. After doing so, we had $N = 8{,}529$ subjects in the full dataset, with $N = 4{,}686$ Whites, $N = 1{,}048$ Blacks, $N = 1{,}729$ Hispanics, and $N = 1{,}039$ in the Other category.

---

[1] R model: *{v} ~ rcs(age) * sex + rcs(height)*, where *v* is each variable.

2.3. Intelligence data

The ABCD dataset contains data from 11 cognitive tests mainly from the NIH toolbox battery (Picture Vocabulary, Flanker, List Sorting, Card Sorting, Pattern Comparison, Picture Sequence Memory, Oral Reading Recognition, the Matrix test from WISC-4, Little Man Test, Rey's auditory learning immediate and delayed recall tests). We adjusted the test data for age and sex, as we were not interested in age and sex differences in this study. After this, the data were imputed using the IRMI algorithm. This is a well-validated deterministic single-imputation approach suitable for reproducible results. The amount of missing data was relatively minor with 10.3% of cells missing, and 48% of cases having at least some missing data. We imputed data for subjects with at most 5 missing datapoints (i.e., at most half the data, rounded down). After the data were imputed, only 1.3% of cells were missing and 98.2% of subjects had no missing data. The subjects who still had missing data were not used in the following analyses.

We extracted the first factor of the 11 neurocognitive tests, which were given at baseline, using standard exploratory factor analysis (EFA) as implemented in the psych package (Revelle, 2020). The general factor accounted for 35% of the variance in test scores, which is lower than the >40% typically found. The smaller portion of variance explained by *g*, in this data, is likely due to the larger number of working memory tests in this set. We used EFA instead of multigroup confirmatory factor analysis, because we did not want to commit to a specific model regarding the nature of cognitive differences between race/ethnic groups (e.g., Spearman's Hypothesis).

2.4. Methods

Because we had a very large number of variables and did not expect strong interactions, we used elastic net as our machine learning algorithm (James et al., 2013). This is a flexible algorithm that is able to deal very well with strong correlations among variables and in training situations where the number of predictors is much greater than the number of cases, as in the current data which had predictor/sample size ratios of 6.3 and 11.5 for the full sample and the White subsample, respectively.

Elastic net has two meta-parameters, one for shrinkage (*penalty*) and one for the form of the shrinkage (*mixture*). The latter decides the relative prominence of the two algorithms, namely ridge regression (with L2 penalization) and the lasso (with L1 penalization), of which the model is a superset. The value is bounded between 0 and 1, with 0 corresponding to the ridge regression and 1 to the lasso. The values in between represent a mix of the two regularization methods. When the mixture value is >0, the geometry of the regularization results in some variables being assigned betas of exactly 0. This method is equivalent to using Bayesian regression with a prior where there is a strong probability of 0 for the betas in the model (i.e., the Laplace prior). This method is appropriate when we think some or most of the predictors have a true beta of 0. Because it was not clear whether variable selection among the 53k variables was needed, we started with the elastic net. We used standard 10-fold cross-validation to estimate the out of sample validity. The methods employed were implemented using the **tidymodels** framework for R (Kuhn et al., 2020). The elastic net engine used was that from **glmnet** (Friedman et al., 2017).

3. Results

3.1. Training on the full sample

We first trained the elastic net algorithm on the full multiracial dataset. The elastic net model results selected a very low amount of variable selection (with a mixture level at or near 0). As such, we subsequently fit the ridge regression model to the same dataset (i.e., we set mixture = 0). Using ridge regression instead of the elastic net algorithm produced marginally better results and, additionally, had the advantage of being less computationally intensive. For this reason, we used ridge regression instead of an elastic net for all subsequent ML analyses.

Figure 1 shows the results for predicting *g* by applying ridge regression to the full sample. The overall model validity was surprisingly good at $r = .61$ and was similar to that based on the elastic net model ($r = .60$). The regression lines did not suggest any predictive validity bias by group, nor were there unexpected nonlinearities.

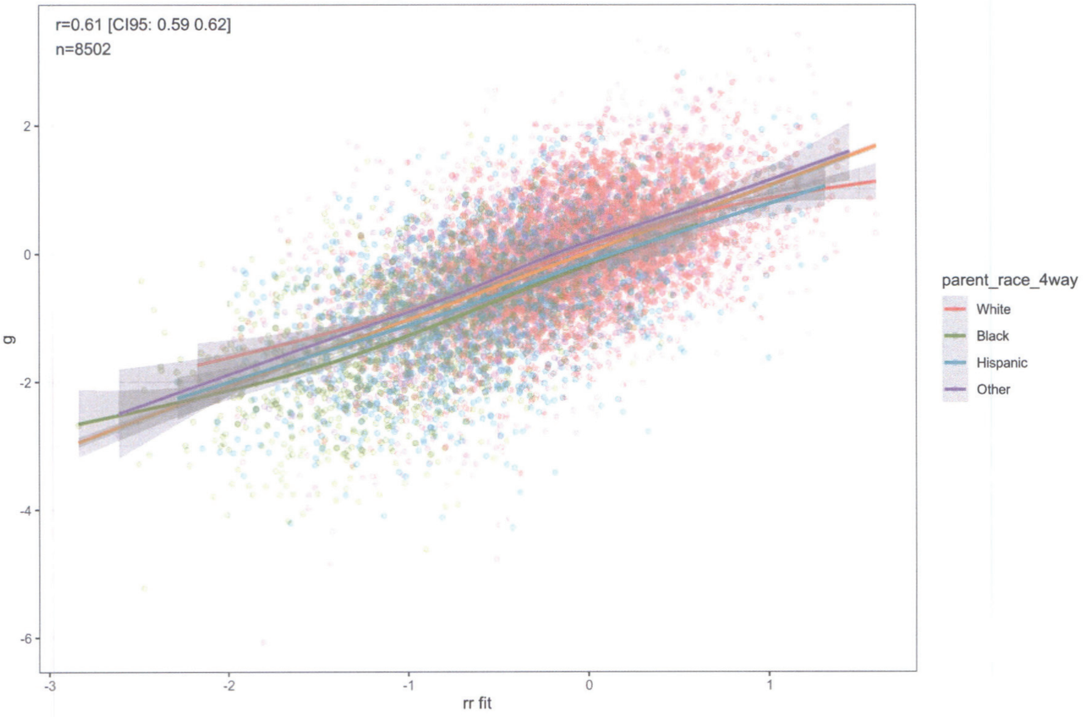


Figure 1. Scatterplot of out of sample predictions for the full dataset based on the ridge regression model. Yellow line = linear fit in the full sample, other lines = nonlinear generalized additive model fit in each subsample

Table 2 shows regression model fits for the out of sample predictions. There are four models to consider. Model 1 provides the baseline prediction of intelligence based on the social race variables. Model 2 provides the baseline prediction based on the MRI-based predictor. It can be seen, when comparing the two, that the MRI-based predictor is far superior to the social race variables when it comes to predicting intelligence. In Model 3, we see that adding race variables did not substantially improve the validity of Model 2. There was only a 1.2% improvement in adjusted R^2. In Model 4, we added interaction terms for the MRI-based predictor and the social race variables. Doing so, gave only a 0.1% adj. R^2 improvement over Model 3.

Table 2. Regression models for intelligence predictions based on the MRI data trained on the full dataset. SR = social race, Predictor = MRI-based predictor * $p<.01$, ** $p<.005$, *** $p<.001$. Unstandardized betas with standard errors in parentheses

Predictor/Model	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Intercept	0.00 (0.015)	0.01 (0.011)	0.07*** (0.013)	0.06*** (0.013)
social race = White	(ref)		(ref)	(ref)
social race = Black	-1.25*** (0.036)		-0.36*** (0.034)	-0.31*** (0.057)
social race = Hispanic	-0.63*** (0.029)		-0.23*** (0.026)	-0.22*** (0.031)
social race = Other	-0.18*** (0.036)		0.05 (0.031)	0.08 (0.033)
MRI-based predictor		1.04*** (0.015)	0.95*** (0.017)	0.91*** (0.023)
Predictor ∗ SR = White				(ref)
Predictor ∗ SR = Black				0.09 (0.053)
Predictor ∗ SR = Hispanic				0.05 (0.043)
Predictor ∗ SR = Other				0.13* (0.049)
R^2 adj.	0.146	0.369	0.381	0.382
N	8502	8529	8502	8502

The results in Model 4 show that the MRI-based predictor largely made social race variables redundant for the purpose of predicting intelligence. This can be interpreted as being due to statistical mediation, which can be seen in the decrease in the size of the betas for race when the MRI-based predictor is included. For instance, the beta for Black went from -1.25 to -0.31, a decrease of 80%. But it could also be due to confounding in the algorithm training.

Furthermore, the interaction estimates for the MRI predictor and social race were all positive. This indicates that the MRI-based predictor worked somewhat better in the non-White subsamples. For the Other race subsample, the interaction was highly significant ($p = .007$). This is perhaps due to this group being very ancestrally heterogeneous and also the algorithm learning to predict genetic ancestry, which is statistically associated with intelligence, from MRI-data (e.g., Kirkegaard et al., 2019).

We next computed the correlations within each social race. These were: rs = .51, .53, .54, and .58, for the White, Black, Hispanic and Other subsamples, respectively. These values are all smaller than the $r = .61$ value for the entire sample. The weaker correlations might have two causes. First, it may be due to confounding whereby the model learns to predict either social race or genetic ancestry in order to predict intelligence through non-neurological pathways. In this case, the validity would be lower within social races because they are more homogeneous in the non-neurological variables.

Second, it might be due to the individual groups having lower variability in intelligence than the combined sample (cf. Table 1), because lower variability in a predictor or a criterion can result in attenuated correlations.

The summary statistics of the predicted MRI scores are shown in Table 3. As seen, the predictions were close to normally distributed in each group as judged by the near-zero values of skew and kurtosis. The variability (*SD*s) of the MRI predictor were somewhat larger for the non-White groups. Strikingly, the means differed substantially by group: Black-White $d = 1.63$, Hispanic-

White $d = 0.73$, Other-White $d = 0.42$. As these gaps are larger than those on the phenotype we were predicting, this suggests confounding.

Table 3. Summary statistics of predicted scores (out of sample). *SD* = standard deviation, MAD = median absolute deviation from the median

Group	***N***	**Mean**	**Median**	***SD***	**MAD**	**Min**	**Max**	**Skew**	**Kurtosis**
White	4686	-0.07	-0.05	0.56	0.56	-2.17	1.60	-0.25	0.09
Black	1048	-1.01	-1.01	0.57	0.60	-2.84	0.61	0.01	-0.27
Hispanic	1729	-0.49	-0.48	0.59	0.60	-2.28	1.31	0.02	-0.10
Other	1039	-0.31	-0.25	0.63	0.64	-2.61	1.45	-0.18	-0.19

3.2 ML prediction of both social race and genetic ancestry from MRI data

To determine if the machine learning model can predict the social race of a person based on MRI data, we fit a multinomial logistic regression model to infer each subject's race. We used the four social race groups (White, Black, Hispanic, and Other) as the outcomes. As before, we used penalization with a mixture of 0, i.e., logistic ridge regression.

We find that the model for predicting social race performs fairly well, with an AUC = .84 and an accuracy of 74.0%. The latter value can be compared to the 55.1% which would result from merely guessing the base rate of the most common group (Whites). As such, the model improvement in terms of reduced error rate was 40% $(1-(100-73.0)/(100-55.1))$.

Table 4 shows the confusion matrix. In the table, we see that the model was able to predict the difference between socially-defined Blacks and Whites with near-perfect accuracy. Of those predicted to be White, 2% were Black; of those predicted to be Black, 2% were White. The results were worse for the other two social race groups, which were ancestrally more heterogeneous. This is not surprising because, from a commonsense hereditarian perspective, brain variation will mostly follow genetic ancestry not social race. Insofar as social race fails to proxy genetic ancestry, the predictions of social race from MRI data is expected to be poorer.

Table 4. Confusion matrix for social race predictions based on MRI data

MRI predicted race	**Social race**			
	White	Black	Hispanic	Other
White	0.77	0.02	0.11	0.10
Black	0.02	0.79	0.08	0.11
Hispanic	0.14	0.06	0.66	0.14
Other	0.15	0.19	0.22	0.44

To examine the issue further, we trained a new model to predict genetic ancestry instead of social race. We limited the analysis to persons with genetic ancestry from the three largest clusters of ancestry in our dataset: European, African, and Amerindian. These genetic ancestry estimates were reused from a prior study (Fuerst et al., 2021). The resulting subset of data had 5,348 subjects with complete data. As ancestries must sum to 1 and cannot be negative, we use Dirichlet regression to analyze the data. Specifically, we used the **Compositional** package in R because it allowed for penalization (Tsagris et al., 2022). The results, again using 10-fold cross-validation, are shown in Table 5.

Table 5. Correlations for MRI-predicted genetic ancestry and genetic ancestry

	Predicted European ancestry	**Predicted African ancestry**	**Predicted Amerindian ancestry**	**European ancestry**	**African ancestry**
Predicted African ancestry	-0.93	1			
Predicted Amerindian ancestry	0.07	-0.44	1		
European ancestry	0.91	-0.85	0.10	1	
African ancestry	-0.87	0.89	-0.28	-0.96	1
Amerindian ancestry	-0.22	-0.04	0.61	-0.24	-0.04

We find that MRI-based predictions of genetic ancestry are very accurate. The three correlations of interest are .91, .89, and .61, for European, African, and Amerindian ancestry, respectively. As expected, the correlations were higher for European and African ancestry than for White and Black social race, respectively. Note, the correlation for Amerindian ancestry was lower because the proportion of Amerindian ancestry was low in this sample at 3%. To repeat the conclusion, we see that machine learning (ML) can identify both social race and genetic ancestry from MRI data with a high degree 340ccuracyy.

3.3. Training in the White subsample

Given the above findings, we were concerned about spurious confounding. In this scenario, ML learns to predict genetic ancestry and then uses genetic ancestry to predict intelligence through non-neurological pathways. This is the MRI-data equivalent of the population stratification problem in genomics (e.g., Lawson et al., 2019). For this reason, we next trained the predictor on only the White subsample (which is 98% European in genetic ancestry). The results are shown in Figure 2.

Among Whites, the correlation between the MRI-based predictor and intelligence is only slightly weaker when training on Whites than when training on the full dataset ($r = .48$ vs. $r = .51$). To examine the influence of training sample size, training sample race (Whites vs. everyone), and test sample race (Whites vs. everyone), we trained and tested the model on Whites and everyone and then compared the results. These are shown in Figure 3.

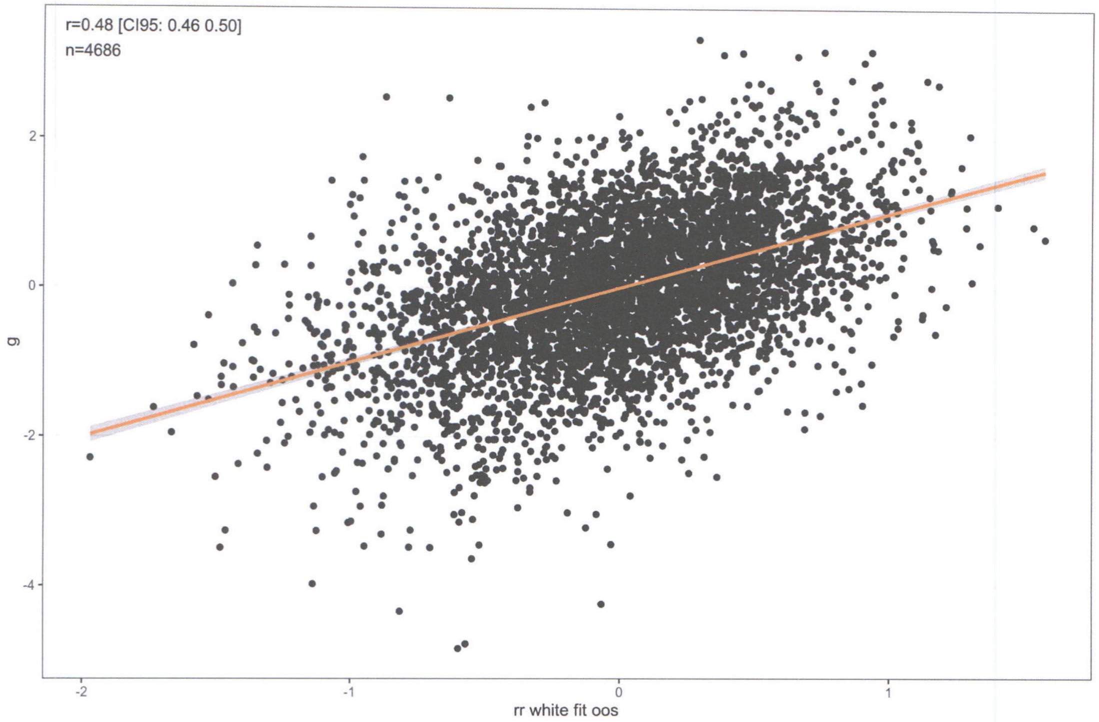


Figure 2. Scatterplot of out of sample predictions of general intelligence (*g*) by the MRI-based predictor for the White subsample based on the ridge regression model

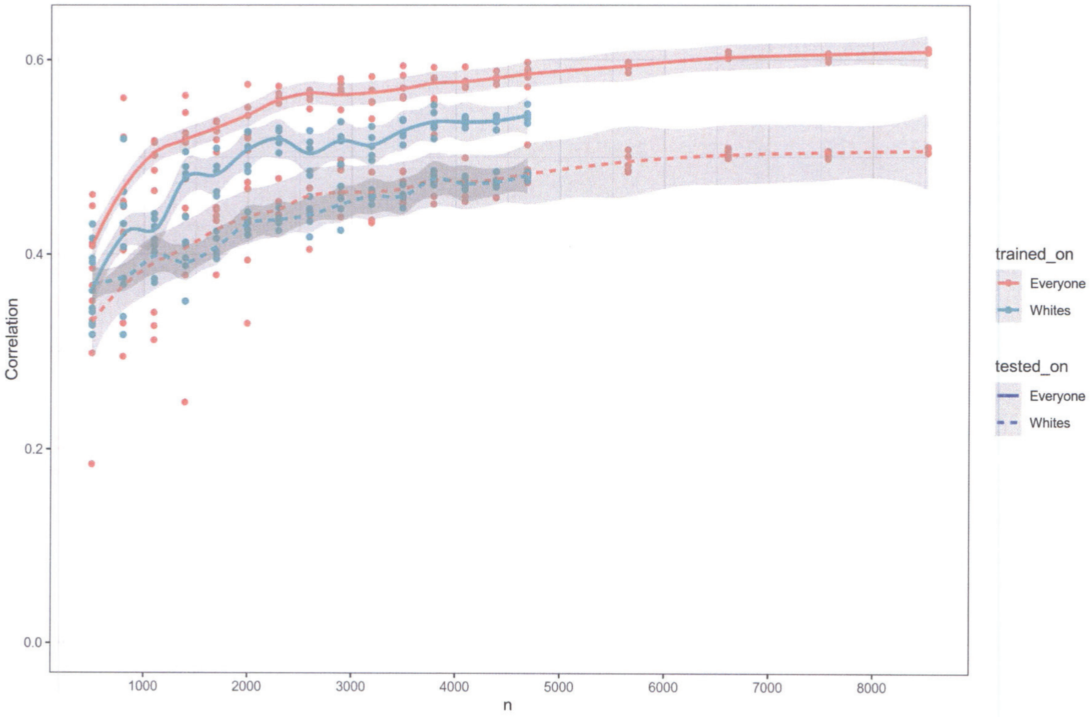


Figure 3. Model predictive validity (10-fold cross-validated correlations) by training sample, test sample, and sample size

There are four lines in Figure 3. By comparing them, one can infer the factors that affect the validity of the MRI-based predictor. First, when comparing the two full lines (in which case the model was tested on everyone) it can be seen that training on Whites vs. everyone reduces the validities. The difference between the two lines reflects the size of the ancestry confounding. Second, when comparing the two turquoise lines (in which case the model was trained on Whites) it can be seen that the testing sample mattered because in this case, the validity is lower in the White subsample. Finally, when comparing the two dotted lines (in which case the model was tested on Whites) it can be seen that the training sample does not matter. Thus, whatever extra information the model is learning when training in the multiracial sample does not affect the validities in the White subsample.

The values obtained for the maximum possible sample size of $N = 4{,}686$ are shown in Table 6. As seen, training and testing on the White subsample decreases the validity by about $r = .10$ as compared to training and testing on the full sample. This decrease cannot be attributed to sample size, as this variable was held constant. The intelligence *SD* is smaller in the White subsample than in the full sample (1.00 vs. 1.13), and this reduced variance may account somewhat for the decrease in the correlations (restriction of range; Hunter & Schmidt, 2015).

Table 6. Model out of sample correlation accuracy by training and test sample at $N = 4{,}686$

	Test set	
Training set	Everyone	Whites
Everyone	0.586	0.487
Whites	0.543	0.481

To further examine the cross-race validity of our model, we scored the entire non-White subsample using the model trained on the White subsample. Because these samples were not used for the model training at all, they cannot be overfitted. For the White sample, we relied on the out-of-sample predictions generated from the cross-validation approach. To maximize the sample size for the White sample, we used 60-fold cross-validation, thus using almost the entire White subsample for training. Figure 4 shows the model predictions across social race. As before, the figure does not reveal any obvious slope interaction by race.

We can evaluate the potential model predictive bias using a regression model. The results are shown in Table 7. As before, there are four models. The first is the baseline race-only model. This shows that without using any MRI data, we can predict 14.6% of the intelligence variance with social race. In contrast, Model 2 shows that by using only the MRI-based predictor, we can predict 29.2% of the variance. Model 3 shows that we get to 34.6% of the variance by using both race and the MRI-based predictor, an increase of 5.4% points. Finally, in Model 4, we add interactions between race and the MRI-based predictor but, as before, doing so does not result in a gain in variance explained.

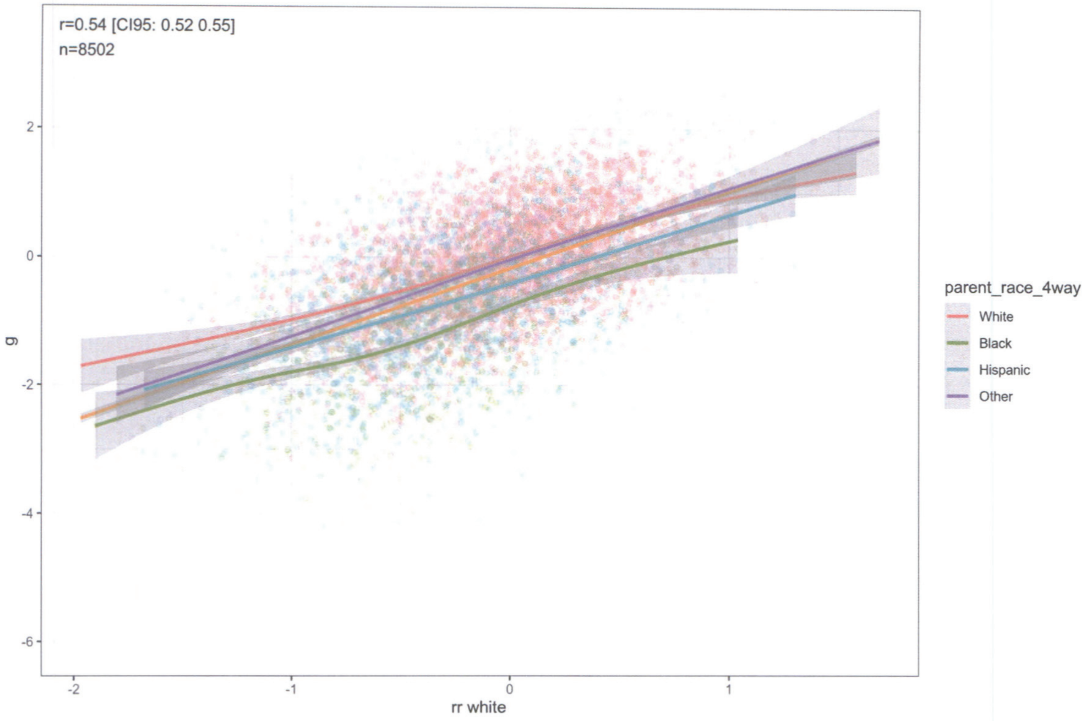


Figure 4. Scatterplot of out-of-sample predictions of general intelligence (*g*) with the MRI-based predictor trained on the White subsample, based on the ridge regression model. Yellow line = linear fit in the full sample, other lines = nonlinear generalized additive model fit in each subsample

Table 7. Regression models for intelligence predictions based on the MRI model trained on the White subsample. SR = social race, Predictor = MRI-based predictor. * $p<.01$, ** $p<.005$, *** $p<.001$. Unstandardized betas with standard errors in parentheses

Predictor/Model	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Intercept	0.00 (0.015)	-0.25*** (0.010)	-0.07*** (0.013)	-0.07*** (0.013)
social race = White	(ref)		(ref)	(ref)
social race = Black	-1.25*** (0.036)		-0.80*** (0.032)	-0.79*** (0.038)
social race = Hispanic	-0.63*** (0.029)		-0.40*** (0.026)	-0.40*** (0.027)
social race = Other	-0.18*** (0.036)		-0.07 (0.031)	-0.07 (0.031)
MRI-based predictor		1.22*** (0.021)	1.06*** (0.021)	1.02*** (0.028)
Predictor * SR = White				(ref)
Predictor * SR = Black				0.05 (0.067)
Predictor * SR = Hispanic				0.03 (0.053)
Predictor * SR = Other				0.17* (0.063)
R^2 adj.	0.146	0.292	0.346	0.346
N	8502	8529	8502	8502

Of the interactions, only the Other social race shows a significant effect ($p < .01$). The slope for the MRI predictor, which is unstandardized, shows estimates around 1.00, indicating this model is well-calibrated. When the MRI-based predictor was standardized, Model 4 showed an effect of $\beta =$ 0.51.

The main contrast with these regression models and those reported in Table 2 (those derived from the MRI models trained on the full dataset) is that the gain from adding social race is larger in these models. Said differently, when trained on Whites, the MRI-based predictor explained less of the effect of race on intelligence. Comparing Model 1 and Model 4 of Table 7 we see, for example, that MRI predictions explain only 37% of the relation between the Black and Hispanic classification and intelligence. This underlines the point we made earlier, namely, that when trained in a multiracial sample, the MRI model will learn to predict race in order to predict intelligence. And when trained only on the White subsample, the model's ability to do this is reduced, so we see a larger role for social race.

As before, we can examine the predicted brain scores to see if they are well behaved and also see how large the race gaps are on these scores. Descriptive statistics are shown in Table 8. There are no notable differences in the dispersion or shape. Moreover, the gaps, while 40% smaller than when the model is trained on the full dataset, are still substantial. In terms of Cohen's *d*, the Black-White, Hispanic-White, and Other-White MRI-based predictor gaps are 0.90*d*, 0.45*d*, and 0.22*d*, respectively. Multiplying these Cohen's *d*s by the validity of the predictor ($\beta = 0.51$; Model 4 in Table 7) gives the approximate size of the intelligence gaps which are explainable by our MRI-based predictor (e.g., Black-White $0.90d * .51\beta = .46d$).

Table 8. Summary statistics of out-of-sample predictions based on the MRI model trained on the White subsample

Social race	*N*	mean	median	SD	MAD	min	max	skew	kurtosis
White	4686	0.00	0.02	0.48	0.49	-1.97	1.58	-0.23	0.01
Black	1048	-0.44	-0.44	0.47	0.48	-1.90	1.04	0.00	-0.05
Hispanic	1729	-0.22	-0.20	0.49	0.49	-1.68	1.30	-0.04	-0.20
Other	1039	-0.11	-0.09	0.51	0.51	-1.80	1.69	-0.08	0.06

3.4. Comparison between the MRI-based predictor and brain volume

We can compare the reported results to those derived from whole brain volume. Table 9 repeats the models of Table 9, replacing the MRI-based predictor with whole brain volume. We see that whole brain volume is a relatively weak predictor in this sample. Model 4 gives the validity for predicting intelligence, which is $\beta = .15$. As such, the MRI-based predictor is 3.4 times better at predicting intelligence. Regarding the size of differences, the Black-White, Hispanic-White, and Other-White whole brain volume differences are, respectively, 0.99*d*, 0.54*d*, and 0.36*d*. Thus, the MRI-based predictor exhibits somewhat smaller gaps than whole brain volume. However, the predictor's validity is much higher and so it explains a larger percentage of group differences.

Table 9. Regression models for intelligence based on whole brain volume. * p<.01, ** p<.005, *** p<.001. SR = social race, Volume = whole brain volume. Standardized betas with standard errors in parentheses

Predictor/Model	Model 1	Model 2	Model 3	Model 4
Intercept	0.00 (0.015, 1)	-0.32*** (0.012)	-0.06*** (0.016)	-0.05** (0.016)
social race = White	(ref)		(ref)	(ref)
social race = Black	-1.25*** (0.036)		-1.07*** (0.037)	-1.10*** (0.044)
social race = Hispanic	-0.63*** (0.029)		-0.54*** (0.030)	-0.54*** (0.030)
social race = Other	-0.18*** (0.036)		-0.12*** (0.036)	-0.13*** (0.036)
Whole brain volume		0.30*** (0.012)	0.17*** (0.012)	0.15*** (0.016)
Volume * SR = White				(ref)
Volume * SR = Black				-0.01 (0.039)
Volume * SR = Hispanic				0.05 (0.032)
Volume * SR = Other				0.10* (0.03)
R^2 adj.	0.146	0.071	0.165	0.165
N	8502	8529	8502	8502

3.5. Relative validities of different sets of MRI variables

We found relatively strong predictive validities for our MRI-based predictor which used data from all available MRI modalities. Here, we explore which modalities contribute most to the predictor's accuracy, and also the relation between these modalities and other variables. To do so, we created MRI-based predictors for each of the 81 datafiles of MRI variables, which are the output from different neuroimaging software. For each dataset, we then computed the number of variables, the validities for intelligence among Whites, the Black-White and Hispanic-White MRI-based predictor gaps, and the correlation between the predictor and polygenic scores for education/intelligence (eduPGS) among Whites. For eduPGS, we used the MTAG eduPGS from Lee et al. (2018). See Fuerst et al. (2021) for details. Table 10 shows the correlation matrix for these variables. Overall, the same MRI-based predictors that are good at predicting intelligence among Whites correlate more strongly with polygenic scores among Whites, exhibit larger racial gaps, and are based on a larger number of variables.

Table 10. Correlation matrix of MRI predictive models resulting from 81 file sets. Values above .30 are $p<.01$

	[1]	[2]	[3]	[4]
[1] Number of variables in dataset	1	0.53	0.28	0.35
[2] Correlation with intelligence (among Whites)	0.53	1	0.68	0.71
[3] Black-White gap	0.28	0.68	1	0.88
[4] Hispanic-White gap	0.35	0.71	0.88	1
[5] Polygenic score correlation (among Whites)	0.37	0.88	0.50	0.61

Figure 5 plots the predictors' validities with respect to intelligence against the predictors' correlations with eduPGS. As can be seen, the top file sets are derived from the n-back task fMRI. Though the scores from the n-back tests were not used to calculate intelligence scores or to derive any of the MRI-based variables, these variables nevertheless are the most predictive. This finding is unexpected based on the reliability of the different MRI data, as sMRI variables are the most reliable while task fMRI variables are the least (Elliott et al., 2020). It is possible that the lack of reliability of each fMRI variable is offset by the large number of them, thus providing a good aggregate signal. It is interesting to note that the values from the first run of n-back are more useful than the second. Perhaps the learning that occurred during the subjects' first exposure to the n-back provided important information about learning speed, while the second run provided less, as the most rapid phase of learning was already over.

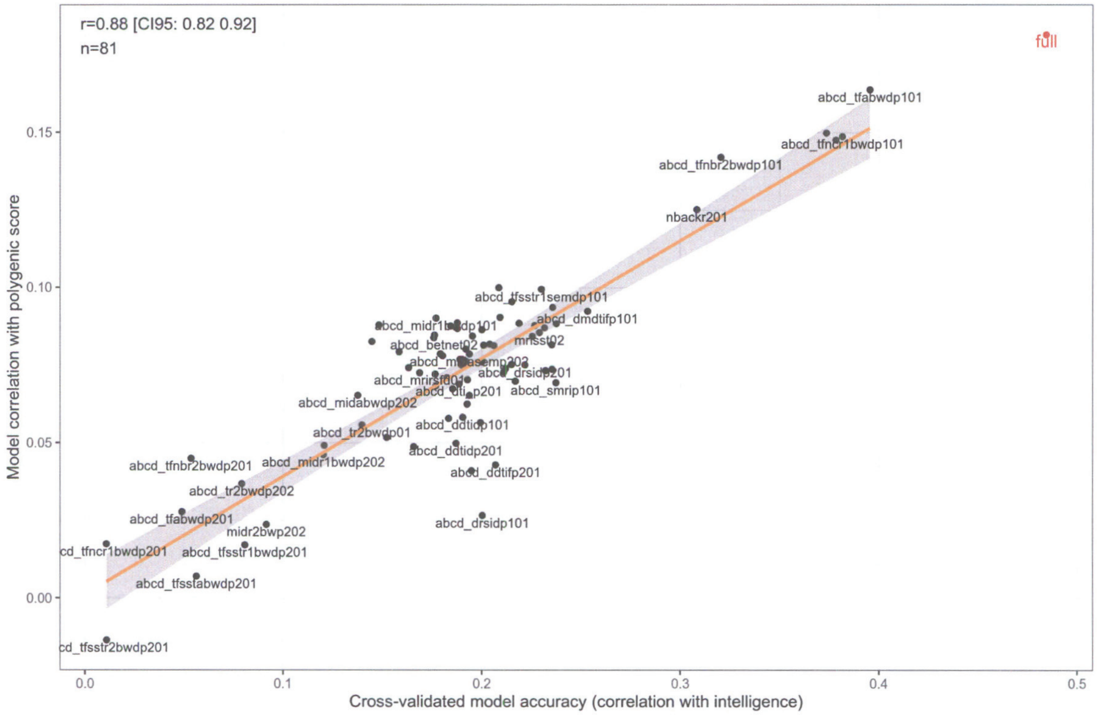


Figure 5. Scatterplot of MRI-based predictor correlation with intelligence and correlation with eduPGS

3.6. Variable Selection

In this section, we summarize the methods used when attempting to reduce the number of predictors in the model without reducing validity. As previously noted, we trained the elastic net model on the complete set of variables. However, the mixture parameter was tilted strongly towards ridge regression and away from the LASSO and, with it, variable selection. For this reason, we used ridge regression in the above analyses. However, given the large number of variables, we decided to make further attempts at variable selection.

The first method we tried was the grouped LASSO (Friedman et al., 2017), as implemented in the **gglasso** package (Yang et al., 2020). Grouped LASSO works the same as the regular LASSO, except that groups of variables are selected together. As we saw in the previous section, on results by file set, some file sets contain much more signal than others for the purpose of predicting intelligence. This suggests that one could select among files instead of among variables. In line with this idea, we used the grouped LASSO to select datasets ($k=81$ datasets). The results from this approach were not much better than using the elastic net on the full data. The best model contained predictors from 47 file sets, totalling 32,569 variables, about 60% of the total.

Because of this unsatisfactory result, we tried a simpler approach. We used grouped forward selection. In forward selection, one begins with an empty model (intercept only). Then each possible predictor is considered and the best ones are chosen by some fit criterion such as BIC. The grouped forward selection works the same way, except with groups of variables. For the criterion, we used the cross-validated correlation, with a minimum increment of .005. Thus, the algorithm continues adding file sets of predictors to the model until the model no longer improves its cross-validated correlation by at least .005.

This algorithm resulted in the selection of 7 file sets totalling 6,070 variables. The cross-validated correlation for this set was $r=.50$, as compared to $r=.48$ when using all variables. As we did not have a holdout sample, we have no way to verify this surprising result, given that this could result from even a small amount overfitting. However, we consider it to be a useful starting point for researchers interested in predicting intelligence using the current dataset. The selected set is:

1. abcd_tfabwdp101 (Beta weights for n-Back task fMRI in cortical (Destrieux) ROIs averaged across runs)
2. abcd_dmdtifp101 (dMRI DTI of fractional anisotropy, part 1)
3. mrisstsem01 (Standard error of the mean for SST task fMRI in subcortical (aseg) and cortical (Desikan aparc) ROIs averaged across all runs)
4. abcd_tfncr1bwdp101 (Beta weights for run 1 n-back task fMRI in cortical (Destrieux) ROIs)
5. abcd_mrisdp101 (sMRI morphometric and image intensity measures, Destrieux ROIs, part 1)
6. abcd_midasemdp101 (Standard error of the mean for MID task fMRI in cortical (Destrieux) ROIs averaged across runs)
7. abcd_mrisdp201 (sMRI morphometric and image intensity measures, Destrieux ROIs, part 2).

Of the selected file sets, there are two n-back task fMRI datasets, two sMRI datasets, one dMRI dataset, and two file sets with standard errors for other variables. The utility of the standard errors in these linear models is surprising and deserves further study.

3.7. Potential family and site confounding

We fit a number of further validation models. We were concerned about family confounding, so we subset the data to the first member from each family. Doing so did not substantially affect results, because the predictors trained on white singletons and trained with siblings also included correlated at $r=.98$. For example, the validity in the subsample of White singletons was $r=.48$, ($N=4528$) as compared to $r=.49$ ($N=4686$) for the sample with siblings. In fact, if it had affected the results, the validity of the White subsample would have been higher, as sometimes members of the same White family would have been split between the different parts of the 10-fold cross-validation resampling. So, we used the models with family members included to maximize sample size.

There is a potential concern regarding the possible effects of the MRI scanning protocol, specifically related to the MRI scanner site. These effects are confounded with socioeconomic factors that may vary between recruitment sites. In response to this concern, we fitted models that controlled for site. However, our analyses revealed that controlling for site did not have an interpretatively significant impact on the results. Therefore, we ultimately used the models without site adjustments to avoid inadvertently controlling for the effects of socioeconomic variables that vary between sites.

4. Discussion

This study found many things of interest. First, we found that a simple machine learning algorithm fit on a large number of MRI variables was able to predict intelligence with high accuracy (β = .51). This was achieved when relying only on an additive, linear model with an appropriate amount of penalization. This high validity contrasts with that of brain volume, which in this dataset was β = .15. The relatively low validity of brain volume is probably due to the low average age (10) of the sample. Because the ABCD is a longitudinal study which will follow the youth until 18 years of age, it will be possible to determine in the near future if the validity of brain volume increases with age.

Second, we found that the predictive validity of the model was inflated when training on a multiracial dataset. This appeared to be because the model learned to predict subjects' social race/genetic ancestry based on the MRI data, and then used this information to predict intelligence. This finding is consistent with those of Gichoya et al. (2022), who report that machine learning can recognize self-reported race/ethnicity from a wide variety of medical imaging data. This is unsurprising because, in the USA, socially identified race closely tracks continental genetic ancestry (Kirkegaard, 2021; Tang et al., 2005), which certainly is not "only skin deep". This race-related confounding has implications for other studies which train on mixed-race/ancestry samples, including a recent special issue of 21 papers which analyzed the ABCD dataset to create brain predictors of fluid intelligence (Pohl et al., 2019).

Nonetheless, when we trained the predictor on the White group, the largest socially identified race in the sample, the model validities remained high and nearly the same across other groups. This mirrors decades-long results based on intelligence tests which find negligible race bias (Brown et al., 1999; Jensen, 1980). Moreover, we find MRI-based predictor differences between social races. These predictor differences can statistically explain about 37% of the relation between both the Black and Hispanic classification and intelligence.

We further analyzed which aspects of MRI data were most useful in the prediction of intelligence. We found that fMRI task data, which measures blood flow when completing tasks, had the highest validity for predicting intelligence. Additionally, MRI datasets which had more variables, which showed larger race differences, and which had higher correlations with polygenic scores, had higher validities.

This study focused on predictive validity, not the question of genetic or environmental causation of MRI-predictor differences. While we found that better MRI predictors of intelligence were more highly associated with polygenic scores for intelligence, it is nonetheless true that brain features are environmentally influenced. Future analyses which are solely interested in brain predictors of heritable intelligence could try to train the MRI-predictor on polygenic scores of intelligence instead of phenotypic intelligence and then examine to what extent these MRI predictors of heritable intelligence relate to phenotypic intelligence and also demographic variables.

We did not utilize a hold-out sample because we already had a quite low number of cases relative to predictors. This would then tend to result in some overfitting insofar as many design choices were made during the study. However, because we used only simple algorithms (additive and linear models without interactions), the risk of substantial overfitting should be very small. Still, it is desirable to validate the predictions from a model trained on ABCD data in another dataset, a true out-of-dataset cross-validation. It is also possible to use the 17% of subjects who did not provide task fMRI data in the present data as a test sample if these provide the necessary data in the follow-up data collection in a few years.

Author contributions: Analyses with individual-level data were conducted by JGRF between 2020 and 2021. EOWK conceptualized the analyses, drafted and consulted on the code, and wrote and edited the manuscript.

Acknowledgments: Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data, but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the research results and interpretations of the authors alone and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Version 3.01. The raw data are available at https://nda.nih.gov/edit_collection.html?id=2573. Additional support for this work was made possible from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680), and Children and Screens: Institute of Digital Media and Child Development Inc. NDA waved the requirement of creating an NDA data project (NDA, Oct 26, 2022).

References

Brown, R.T., Reynolds, C.R. & Whitaker, J.S. (1999). Bias in mental testing since *Bias in Mental Testing*. *School Psychology Quarterly* 14(3): 208-238.

Brueggeman, L., Koomar, T., Huang, Y., Hoskins, B., Tong, T., Kent, J., …, & Michaelson, J.J. (2019). Ensemble modeling of neurocognitive performance using MRI-derived brain structure volumes. In: K.M. Pohl, W.K. Thompson, E. Adeli & M.G. Linguraru (eds.), *Adolescent Brain Cognitive Development Neurocognitive Prediction*, pp. 124-132. Springer International.

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., … & Dale, A.M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience* 32: 43-54.

Cox, S.R., Ritchie, S.J., Fawns-Ritchie, C., Tucker-Drob, E.M. & Deary, I.J. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence* 76: 101376.

DeSilva, J.M., Traniello, J.F.A., Claxton, A.G. & Fannin, L.D. (2021). When and why did human brains decrease in size? A new change-point analysis and insights from brain evolution in ants. *Frontiers in Ecology and Evolution* 9: 742639.

Elliott, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., ... & Hariri, A.R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science* 31: 792-806.

Farias, S.T., Mungas, D., Reed, B., Carmichael, O., Beckett, L., Harvey, D., ... & DeCarli, C. (2012). Maximal brain size remains an important predictor of cognition in old age, independent of current brain pathology. *Neurobiology of Aging* 33: 1758-1768.

Friedman, J., Hastie, T., Simon, N., Qian, J. & Tibshirani, R. (2017). *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models* (2.0-13) [Computer software]. https://cran.r-project.org/web/packages/glmnet/index.html

Fuerst, J. (2015). Nature of race: The genealogy of the concept and the biological construct's contemporaneous utility. *Open Behavior Genetics*.

Fuerst, J.G.R., Hu, M. & Connor, G. (2021). Genetic ancestry and general cognitive ability in a sample of American youths. *Mankind Quarterly* 62: 186-216.

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R.Z., Heeringa, S., ... & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience* 32: 16-22.

Gichoya, J.W., Banerjee, I., Bhimireddy, A.R., Burns, J.L., Celi, L.A., Chen, L.C., ... & Zhang, H. (2022). AI recognition of patient race in medical imaging: A modelling study. *Lancet Digital Health* 4(6): e406-e414.

Gignac, G.E. & Bates, T.C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence* 64:18-29.

Hagler, D.J., Hatton, S.N., Cornejo, M.D., Makowski, C., Fair, D.A., Dick, A.S., … & Dale, A.M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage* 202: 116091.

Haier, R.J. (2016). *The Neuroscience of Intelligence*, 1st edition. Cambridge Univ. Press.

Hunter, J.E. & Schmidt, F.L. (2015). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 3rd edition. SAGE.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (eds.) (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer.

Jensen, A R. (1980). *Bias in Mental Testing*. Free Press.

Jensen, A.R. & Johnson, F.W. (1994). Race and sex differences in head size and IQ. *Intelligence* 18: 309-333.

Jernigan, T.L., Brown, S.A. & Dowling, G.J. (2018). The Adolescent Brain Cognitive Development Study. *Journal of Research on Adolescence* 28(1): 154-156.

Jiang, R., Calhoun, V.D., Cui, Y., Qi, S., Zhuo, C., Li, J., …, & Sui, J. (2020a). Multimodal data revealed different neurobiological correlates of intelligence between males and females. *Brain Imaging and Behavior* 14: 1979-1993.

Jiang, R., Calhoun, V.D., Fan, L., Zuo, N., Jung, R., Qi, S., … & Sui, J. (2020b). Gender differences in connectome-based predictions of individualized intelligence quotient and sub-domain scores. *Cerebral Cortex* 30: 888-900.

Kievit, R.A., van Rooijen, H., Wicherts, J.M., Waldorp, L.J., Kan, K.-J., Scholte, H.S. & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience* 3(2): 89–97.

Kirkegaard, E.O.W. (2021). Genetic ancestry and social race are nearly interchangeable. *OpenPsych*.

Kirkegaard, E., Woodley of Menie, M., Williams, R., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych* 1: 1-25.

Kuhn, M., Wickham, H. & RStudio (2020). *tidymodels: Easily Install and Load the "Tidymodels" Packages* (0.1.0) [Computer software]. https://CRAN.R-project.org/package=tidymodels

Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., … & Timpson, N.J. (2019). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* 139: 23-41.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & 23andMe Research Team (2018). Gene discovery and polygenic prediction from a 1.1-million person GWAS of educational attainment. *Nature Genetics* 50: 1112-1121.

Li, T., Wang, X., Luo, T., Yang, Y., Zhao, B., Yang, L., Zhu, Z. & Zhu, H. (2019). Adolescent fluid intelligence prediction from regional brain volumes and cortical curvatures using BlockPC-XGBoost. In: K.M. Pohl, W.K. Thompson, E. Adeli & M.G. Linguraru (eds.), *Adolescent Brain Cognitive Development Neurocognitive Prediction*, pp. 167-175. Springer International.

Office of Management and Budget (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register* 62(210): 58782-58790.

Pietschnig, J., Gerdesmann, D., Zeiler, M. & Voracek, M. (2022). Of differing methods, disputed estimates and discordant interpretations: The meta-analytical multiverse of brain volume and IQ associations. *Royal Society Open Science* 9(5): 211621.

Pohl, K.M., Thompson, W.K., Adeli, E. & Linguraru, M.G. (2019). *Adolescent Brain Cognitive Development Neurocognitive Prediction*. Springer.

Ren, H., Wang, X., Wang, S. & Zhang, Z. (2019). Predict fluid intelligence of adolescent using ensemble learning. In: K.M. Pohl, W.K. Thompson, E. Adeli & M.G. Linguraru (eds.), *Adolescent Brain Cognitive Development Neurocognitive Prediction*, pp. 66-73. Springer International.

Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research* (1.9.12.31) [Computer software]. https://CRAN.R-project.org/package=psych

Rushton, J.P. & Ankney, C.D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience* 119: 692-732.

Sesardic, N. (2010). Race: A social destruction of a biological concept. *Biology & Philosophy* 25(2): 143-162.

Sripada, C., Angstadt, M. & Rutherford, S. (2018). Towards a "treadmill test" for cognition: Reliable prediction of intelligence from whole-brain task activation patterns. *BioRxiv*: 412056.

Tang, H., Quertermous, T., Rodriguez, B., Kardia, S.L., Zhu, X., Brown, A., ... & Risch, N.J. (2005). Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *American Journal of Human Genetics* 76: 268-275.

Tsagris, M., Athineou, G., Alenazi, A. & Adam, C. (2022). *Compositional: Compositional Data Analysis* (5.7) [Computer software]. https://CRAN.R-project.org/package=Compositional

Valverde, J.M., Imani, V., Lewis, J.D. & Tohka, J. (2019). Predicting intelligence based on cortical WM/GM contrast, cortical thickness and volumetry. In: K.M. Pohl, W.K. Thompson, E. Adeli & M.G. Linguraru (eds.), *Adolescent Brain Cognitive Development Neurocognitive Prediction*, pp. 57-65. Springer International.

Wang, L., Wee, C.-Y., Suk, H.-I., Tang, X. & Shen, D. (2015). MRI-based intelligence quotient (IQ) estimation with sparse learning. *PLOS ONE* 10(3): e0117295.

Wlaszczyk, A., Kaminska, A., Pietraszek, A., Dabrowski, J., Pawlak, M.A. & Nowicka, H. (2019). Predicting fluid intelligence from structural MRI using random forest regression. In: K.M. Pohl, W.K. Thompson, E. Adeli & M.G. Linguraru (eds.), *Adolescent Brain Cognitive Development Neurocognitive Prediction*, pp. 83-91. Springer International.

Yang, Y., Zou, H. & Bhatnagar, S. (2020). *gglasso: Group Lasso Penalized Learning Using a Unified BMD Algorithm* (1.5) [Computer software]. https://CRAN.R-project.org/package=gglasso

Zahodne, L.B., Manly, J.J., Narkhede, A., Griffith, E.Y., DeCarli, C., Schupf, N.S., Mayeux, R. & Brickman, A.M. (2015). Structural MRI predictors of late-life cognition differ across African Americans, Hispanics, and Whites. *Current Alzheimer Research* 12: 632-639.

Zhang-James, Y., Glatt, S.J. & Faraone, S.V. (2019). Nu support vector machine in prediction of fluid intelligence using MRI data. In: K.M. Pohl, W.K. Thompson, E. Adeli & M.G. Linguraru (eds.), *Adolescent Brain Cognitive Development Neurocognitive Prediction*, pp. 92-98. Springer International.

SUPPLEMENTAL MATERIALS

Supplemental materials for Chapter 4: Continental genetic ancestry source correlates with global cognitive ability score are available at Open Science Frame data repository: https://osf.io/bc5fv/

Supplemental materials for Chapter 5: Biogeographic ancestry, cognitive ability and socioeconomic outcomes; Chapter 6: Global ancestry and cognitive ability; Chapter 7: More research needed: There is a robust causal vs. confounding problem for intelligence-associated polygenic scores in context to admixed American populations; Chapter 8: Linear and partially linear models of behavioral trait variation using admixture regression; Chapter 9: Genetic ancestry and general cognitive ability in a sample of American youths; Chapter 10: A genetically-informed test of the cognitive-colorism hypothesis; Chapter 12: A genetic hypothesis for American race/ethnic differences in mean g: A reply to Warne (2021) with fifteen new empirical tests using the ABCD dataset; and Chapter 13: A multimodal MRI-based predictor of intelligence and its relation to race/ethnicity are all available at Open Science Frame data repository: https://osf.io/kgavp/

CONTRIBUTORS

Noah Carl is a British sociologist and intelligence researcher. See https://www.researchgate.net/profile/Noah-Carl-2 for a list of his publications.

Nathan Cofnas works in the philosophy of biology and ethics and the intersection of these fields. See https://www.researchgate.net/profile/Nathan-Cofnas for a list of his publications.

Gregory Connor is an econometrician with research interests in the theory and applications of factor modeling. See https://www.researchgate.net/profile/Gregory-Connor for a list of his publications.

John G.R. Fuerst works in experimental psychology and its intersection with genomics. See https://www.researchgate.net/profile/John-Fuerst for a list of his publications.

Meng Hu does research in differential psychology with a focus on intelligence. See https://menghu.substack.com/ for a list of his publications.

Emil O.W. Kirkegaard does quantitative research across a range of social sciences including psychology, demographics, and economic development. See https://www.researchgate.net/profile/Emil-O-W-Kirkegaard for a list of his publications.

Jordan Lasker performs research in the fields of bioinformatics, genetics, behavioural science, and psychology. See https://www.researchgate.net/profile/Jordan-Lasker for a list of his publications.

Gerhard Meisenberg has research interests in psychology, physiology, and biochemistry. *See* https://www.researchgate.net/profile/Gerhard-Meisenberg-2 *for a list of his publications.*

Brian J. Pesta has a range of research interests in differential psychology. See https://www.researchgate.net/profile/Bryan-Pesta for a list of his publications.

Davide Piffer does research in psychometrics, quantitative psychology, and social psychology. See https://www.researchgate.net/profile/Davide-Piffer-2 for a list of his publications.

Vladimir Shibaev performs research in differential psychology with a focus on intelligence. See https://www.researchgate.net/profile/Vladimir-Shibaev for a list of his publications.

Russell T. Warne is a psychometrician with research interests focussed on quantitative psychology. See https://russellwarne.com/ for a list of his publications.

Robert L. Williams works in experimental psychology. See https://www.researchgate.net/profile/Robert-Williams-45 for a list of his publications.

Cambrid
Scholars
Publishin

This book is an edited collection of recently published papers on the sources of average test score gaps when analysed through the lenses of race and ethnicity, socio-economic status, and biogeographic ancestries such as European, African, and East Asian ancestry. It brings together exciting recent findings that rely on powerful DNA-based methods developed in the last few decades. The book also considers the public policy question as to whether, and how, these findings should be disseminated to the general public audience.

"This anthology provides an excellent overview of the research on test score gaps and their empirical relationship with and possible theoretical explanation through genetic factors. Widely scattered individual works are brought together in one volume and presented with an introductory text that is understandable to a broader public. A must-read for everyone interested in the topic and especially for those who are initially skeptical about this research but want to get to know it better for a fruitful scientific discussion."
Dr Heiner Rindermannn
***Professor of Educational and Developmental Psychology at Chemnitz University of Technology, Germany, and author of* Cognitive Capitalism: Human Capital and the Wellbeing of Nations**

"If you are a quantitative social scientist, you have a responsibility to wrestle with the material in this book. It is important that you do it partly because the genetics of differences in test scores is of the utmost relevance to a wide range of topics involving both scholarly issues and public policy. There is also a larger reason. It is important that you treat its arguments as legitimate scientific inquiry to help rescue social science from the dark age of censorship and ideological orthodoxy into which it has fallen."
Charles Murray
***F.A. Hayek Emeritus Scholar at the American Enterprise Institute, and co-author (with Richard Herrnstein) of* The Bell Curve.**

Gregory Connor recently retired from his position as Professor of Finance at Maynooth University (Ireland). He previously taught at the London School of Economics (UK), the University of California-Berkeley (USA), and Northwestern University (USA). He has published widely and is co-author (with Robert Korajczyk and Lisa Goldberg) of *Financial Risk Analysis*. His research focusses on factor modelling with large cross-sections applied to macro-finance, genome-wide association studies (GWAS), the g-factor, and gene/environment analysis of variance (ANOVA).

John Fuerst has previously conducted research at Cleveland State University (USA), North Carolina State University (USA), and Case Western Reserve University (USA). He is currently enrolled in a post graduate programme in bioinformatics and has previously done graduate work in psychology. His research focusses on behavioural genetics, psychometrics, genome-wide association studies (GWAS), and differential psychology.

978-1-5275-7566-0
www.cambridgescholars.com
Cover image created with the assistance of DALL-E 2
© John Fuerst, 2023

9 781527 575660